[illegible] COLLEGE
LIBRARY
DETROIT, MICH.

JULY/AUGUST 2010 • VOLUME 81, NUMBER 4

CHILD DEVELOPMENT

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

SPECIAL SECTION
Disasters and Their Impact on Child Development

Special Section Editors
ANN S. MASTEN • JOY D. OSOFSKY

Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www3.interscience.wiley.com/aboutus/journal_ordering_and_payment.html#Tax. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/journal-info.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.interscience.wiley.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from csjournals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to interscience.wiley.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at http://www.interscience.wiley.com/support
Visit our Online Customer Self-Help available in 6 languages at http://www.interscience.wiley.com/support

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Detra N. Davis

Editorial Assistant

Antonella Caiazza

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Ramesh Bhatt
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Cara Cashon
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Ana Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Greg Reynolds
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, July/August 2010, Volume 81, Number 4, Pages 1025–1340

Contents

IN THIS ISSUE 1025

SPECIAL SECTION

Disasters and Their Impact on Child Development
Editors: Ann S. Masten and Joy D. Osofsky

INTRODUCTION

Disasters and Their Impact on Child Development: Introduction to the Special Section
Ann S. Masten and Joy D. Osofsky 1029

SPECIAL SECTION EMPIRICAL ARTICLES

Disasters, Victimization, and Children's Mental Health
Kathryn A. Becker-Blease, Heather A. Turner, and David Finkelhor 1040

Unpacking Trauma Exposure Risk Factors and Differential Pathways of Influence: Predicting Postwar Mental Distress in Bosnian Adolescents
Christopher M. Layne, Joseph A. Olsen, Aaron Baker, John-Paul Legerski, Brian Isakson, Alma Pašalić, Elvira Duraković-Belko, Nermin Đapo, Nihada Ćampara, Berina Arslanagić, William R. Saltzman, and Robert S. Pynoos 1053

Sierra Leone's Former Child Soldiers: A Follow-Up Study of Psychosocial Adjustment and Community Reintegration
Theresa Stichick Betancourt, Ivelina Ivanova Borisova, Timothy Philip Williams, Robert T. Brennan, Theodore H. Whitfield, Marie de la Soudiere, John Williamson, and Stephen E. Gilman 1077

Posttraumatic Resilience in Former Ugandan Child Soldiers
Fionna Klasen, Gabriele Oettingen, Judith Daniels, Manuela Post, Catrin Hoyer, and Hubertus Adam 1096

The Kenyan Political Conflict and Children's Adjustment
Mumbe Kithakye, Amanda Sheffield Morris, Andrew M. Terranova, and Sonya S. Myers 1114

Impact of Maternal Posttraumatic Stress Disorder and Depression Following Exposure to the September 11 Attacks on Preschool Children's Behavior
Claude M. Chemtob, Yoko Nomura, Khushmand Rajendran, Rachel Yehuda, Deena Schwartz, and Robert Abramovitz 1129

Exposure to 9/11 Among Youth and Their Mothers in New York City: Enduring Associations With Mental Health and Sociopolitical Attitudes
Elizabeth T. Gershoff, J. Lawrence Aber, Angelica Ware, and Jennifer A. Kotler 1142

The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories
Elif Celebi Oncu and Aysegul Metindogan Wise 1161

Tsunami, War, and Cumulative Risk in the Lives of Sri Lankan Schoolchildren
Claudia Catani, Abigail H. Gewirtz, Elizabeth Wieling, Elisabeth Schauer, Thomas Elbert, and Frank Neuner 1176

Growing Pains: The Impact of Disaster-Related and Daily Stressors on the Psychological and Psychosocial Functioning of Youth in Sri Lanka
Gaithri A. Fernando, Kenneth E. Miller, and Dale E. Berger 1192

Exploring Posttraumatic Growth in Children Impacted by Hurricane Katrina: Correlates of the Phenomenon and Developmental Considerations
Ryan P. Kilmer and Virginia Gil-Rivas 1211

Sex Differences in Salivary Cortisol, Alpha-Amylase, and Psychological Functioning Following Hurricane Katrina
Jacob M. Vigil, David C. Geary, Douglas A. Granger, and Mark V. Flinn 1228

Children of Katrina: Lessons Learned About Postdisaster Symptoms and Recovery Patterns
Mindy E. Kronenberg, Tonya Cross Hansel, Adrianne M. Brennan, Howard J. Osofsky, Joy D. Osofsky, and Beverly Lawrason 1241

Children With Disabilities in the Context of Disaster: A Social Vulnerability Perspective
Lori Peek and Laura M. Stough 1260

Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study
Alastair Ager, Lindsay Stark, Bree Akesson, and Neil Boothby 1271

EMPIRICAL ARTICLES

Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations
Erik D. Thiessen and Meagan N. Yee 1287

An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children
Petra H. J. M. Vlamings, Lisa M. Jonkman, and Chantal Kemner 1304

Typologies of Family Functioning and Children's Adjustment During the Early School Years
Melissa L. Sturge-Apple, Patrick T. Davies, and E. Mark Cummings 1320

OTHER

Manuscripts Accepted 1336

Child Development, July/August 2010, Volume 81, Number 4, Pages 1025–1028

In This Issue

Effects of Disaster and Victimization

In a study of more than 2,000 American children ages 2–17, **Becker-Blease, Turner, and Finkelhor** (p. 1040) looked at the effects of victimization (human-caused harm such as maltreatment, abuse and neglect, and assault) and disaster (one-time or ongoing events such as hurricanes and earthquakes that produce stress). Almost 14% of the children had experienced lifetime exposure to disaster, whereas more than 4% had been involved in a disaster in the past year. Children who were exposed to victimization only or to both disaster and victimization had higher levels of anxiety, depression, and aggression than those who were not exposed. By providing the first prevalence and incidence estimate of disasters in a nationally representative sample of U.S. children, the study helps us understand the interactions among various kinds of stressful events.

Researchers who work in areas where people have been harmed by disasters often focus solely on the impact of direct exposure to the disaster. **Fernando, Miller, and Berger** (p. 1192), in their study of tsunami survivors in Sri Lanka, argue that it is important to consider the role of daily stressors as well. Looking at more than 400 Sri Lankan youths ages 11–20, they find that while war and disaster had a direct effect on youths' psychological health, daily stressors—including poverty, family violence, and lack of safe housing—represented major sources of ongoing stress. By making sure not to miss the importance of ongoing stressors in youths' daily lives, the study highlights the need for holistic policies and programs that focus on reducing these current stressors.

Millions of children are affected by man-made disasters such as having their homes destroyed or losing a parent, but most research focuses on survival and physical well-being, overlooking the social and emotional impacts. In a longitudinal study, **Kithakye, Morris, Terranova, and Myers** (p. 1114) looked at these factors among more than eighty 3- to 7-year-old children who experienced political conflict in Kenya; data were collected before and after the conflict. As expected, children with worse conflict experiences were more likely to have more social and emotional difficulties, specifically more aggressive behavior and less positive, prosocial behavior. Children who were able to adapt their emotional responses and control their emotions were better able to cope with the conflict. The findings have implications for those working to help child victims of conflict.

Experiences From 9/11

Chemtob et al. (p. 1129) took a longitudinal look at preschoolers exposed to the September 11 attacks in New York City. Were children ages 5 or younger whose mothers suffered from posttraumatic stress disorder (PTSD) and depression as a result of the terrorist attacks more likely to have behavior problems than children of moms with neither PTSD nor depression, or with only depression? In their 21-month study of more than 100 moms and their youngsters who were directly exposed to the attacks, Chemtob et al. find that moms who had PTSD and were depressed (compared to those who were only depressed or had neither disorder) were much less likely to be able to help their children regulate their emotions and behavior. As a result, the children were more likely to have problems such as aggression, anxiety, depression, and difficulty sleeping. In suggesting that very young children are highly sensitive to their parents, the findings point to the value of identifying at-risk youngsters through their mothers' experiences.

Did exposure to the 9/11 attacks among adolescents and moms living in New York City result in lasting mental health problems such as posttraumatic stress syndrome (PTSD) and depression? Did it affect social attitudes a year after the attacks? **Gershoff, Aber, Ware, and Kotler** (p. 1142) studied more than four hundred 12- to 20-year-olds and their moms. They were surprised to find that a year after 9/11, exposure to the terrorist attacks did not have a significant effect on the youths' and their mothers' mental health. However, there were elevated levels of problematic mental health symptoms among the mother–child pairs, suggesting that something other than the trauma of 9/11—perhaps everyday stress and trauma—is involved. The

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0001

study also finds that exposure to 9/11 appears to have prompted families to have more discussions about serious social and political issues, including mistrust and prejudice toward immigrants.

Child Exposure to Wars and Disasters

Children are the most vulnerable group for harmful long-term psychological consequences of violence and destruction. In Sri Lanka, residents living in the northeast have been affected by a civil war as well as the 2005 tsunami. **Catani et al.** (p. 1176) sought to determine what happens to children when various risk factors pile up. In their study, children in the resource-poor country were affected by both armed conflict and a natural disaster, and many also coped with domestic violence. The epidemiological study, of almost 1,400 Tamil children ages 9–15, used data from four cross-sectional studies. Findings: All of the adverse experiences contributed significantly to children's difficulties adapting. Particularly stressful were very severe exposure to trauma and loss of family support through violence or death. The findings have implications for efforts to protect traumatized children in developing countries.

Traumatic events that occur during wars and disasters create widespread mental health problems for young people and their families that often outstrip available mental health resources. Currently, clinicians cannot predict which youths will recover on their own and which will require intervention in order to recover. **Layne et al.** (p. 1053) studied 880 secondary-school students (ages 13–19) in Bosnia 5 years after the end of the Bosnian civil war, measuring the types of war trauma to which the students had been exposed. Then they identified the types of exposure that were most likely to lead to mental distress. By developing a way to identify which high-risk youths and their families are most likely to need mental health services, the study provides tools for clinicians involved in helping war-exposed youths.

Child Soldiers

Despite international bans, more than 250,000 children are exploited as soldiers worldwide, almost half of them in Africa. In a study of former Ugandan child soldiers, **Klasen et al.** (p. 1096) looked at 330 children ages 11–17. Most were exposed to shooting and beaten by armed forces, half had killed, and a quarter had been raped. A third suffered from PTSD, more than a third were depressed, and more than half had behavioral and emotional problems. Despite their exposure to severe trauma, however, a third of the former child soldiers exhibited posttraumatic resilience as indicated by the absence of clinically significant mental health problems. Why? Once they returned from war, these children had less exposure to violence in their homes and communities, fewer feelings of survivor guilt, less motivation to seek revenge, better socioeconomic situations, and more perceived spiritual support, according to the study. The findings have implications for developing policies and programs for the two thirds of child soldiers who continued to be exposed to violence once they returned home and suffered from significant mental health problems.

In separate research, the first longitudinal study of psychosocial adjustment in former child soldiers involved in Sierra Leone's bloody civil war, **Betancourt et al.** (p. 1077) looked at more than 150 children ages 10 and 18, mostly male, over 2 years. They find that children who injured or killed others in war were more hostile than children who had not done so and that those who survived rape were more anxious and hostile, suggesting that these categories of war trauma are highly toxic to children's psychosocial adjustment. They also find that children who were abducted at a young age were more depressed than those taken later. Children living in communities in which they felt accepted were less depressed and more confident, and children who stayed in school had better attitudes and behaviors. For the children in at least 86 countries who have been abducted and forcefully conscripted into armed forces, this study has implications for long-term treatment.

Earthquake After-Effects

In 1999, two major earthquakes shook provinces in Turkey. To explore the long-term effects of the disasters, 2 years after the natural disaster, **Celebi Oncu and Wise** (p. 1161) studied more than fifty 7-year-old children who had been affected by the earthquakes; the children were asked to complete stories that involved stressful situations. A control group of 50 children who did not experience the earthquakes also completed stories. Children who had been through the earthquakes provided more negative story endings than their peers, especially when they felt they could not change the course of events, suggesting that the effects of trauma are long-lasting.

In the Aftermath of Katrina

Hurricane Katrina struck the Gulf Coast in August 2005, affecting countless children and families. Recovery from the storm was slow, and many children experienced ongoing stressors that resulted in a range of trauma responses. Three studies in this issue address Katrina survivors.

In a longitudinal study of recovery patterns of children affected by Katrina, **Kronenberg et al.** (p. 1241) find that most of the almost 400 elementary- and high-school-aged children they surveyed did not display long-term psychological problems. For most, trauma symptoms such as feeling depressed, sad, nervous, as well as having trouble concentrating and sleeping, declined over time. Moreover, children and adolescents showed enormous strengths, largely, the study finds, because of supportive relationships and rebuilt schools. But this work in progress finds that more than a quarter of the children continued to have significant trauma symptoms of posttraumatic stress and depression 3 years later. Children who reported these symptoms were more likely to be younger and female, have seen a mental health professional, and have reported family or school problems. Continued study of these children will provide further insights into resiliency.

Kilmer and Gil-Rivas (p. 1211) explored what qualities of children and their relationships with their primary caregivers relate to posttraumatic growth—the positive changes that occur as a result of struggles with trauma. In their longitudinal study, they interviewed more than fifty 7- to 10-year-olds affected by Hurricane Katrina, asking about children's views of themselves, how their thoughts influenced their reactions to the trauma, and the role of their caregivers after the hurricane. Findings: Certain types of thoughts and feelings can lead to posttraumatic growth; in particular, both negative, recurrent, distressing thoughts and constructive, repetitive thinking play a role, suggesting that discomfort combined with reflective processes can lead to positive change. Furthermore, while neither caregivers' warmth nor their distress affected growth, caregivers' coping advice encouraging children to look for the positive did. The study sheds light on our understanding of posttraumatic growth in children and can help those working with youth after trauma.

Vigil, Geary, Granger, and Flinn (p. 1228) studied the effects of Katrina from the standpoint of gender. Experiencing distress is a normal part of life, yet males and females have been found to show different patterns of stress reactions in laboratory studies. In their study of more than 60 teens displaced by Hurricane Katrina and living in a government relocation camp, the researchers measured psychological functioning (such as depression and aggression) and two hormone markers of stress regulation (salivary cortisol and α-amylase), then compared these with more than 50 adolescents who were demographically similar but lived in a different part of the United States that was not affected by Katrina. Vigil et al. find that male and female teens who experienced the hurricane showed distinct patterns of behavioral stress as well as gender differences in how they regulated physiological stress. Specifically, compared to the children who were not exposed to Katrina, girls were more depressed and boys exhibited lower levels of aggression and higher levels of confidence. In suggesting that males and females developed specialized systems for processing and responding to stress, the study adds to our understanding of stress regulation following trauma.

Disaster and Disabilities

Disasters affect nearly 70 million children worldwide each year; about a 10th of those children have disabilities, and more acquire disabilities as a result of disaster. In a review of research on how individuals with disabilities cope with disaster and how children are affected by disaster, **Peek and Stough** (p. 1260) find the following: Children with disabilities are more likely to be poor and live in low-quality housing, boosting their exposure to hazards. When disaster strikes, children with disabilities may have a more difficult time taking action, escaping from, or withstanding the force of the event. Even when evacuation occurs successfully, children with disabilities and their families may be less likely to leave the threatened area. Children with disabilities also encounter barriers during emergency response and recovery phases. And they face special challenges postdisaster when separated from their parents, teachers, or other support professionals. This commentary can help inform disaster-preparedness and response efforts for children with disabilities and their caregivers.

Best Practices

How should children be cared for and protected in humanitarian crisis situations? **Ager, Stark, Akesson, and Boothby** (p. 1271) asked 30 specialists from leading organizations responsible for the care and protection of children in humanitarian crises worldwide to identify best practices to inform and improve future programs for children. Consensus was reached on 55 statements, including the need to coordinate among agencies, involve the community in promoting child well-being and development, and be culturally sensitive to different views of childhood in different countries. Responses suggested that two ideas from child development research—resilience and social ecology—have most influenced current views of best practices, but the researchers suggest that findings from developmental science needed to be more clearly communicated to practitioners and policy makers so they can be used in programming.

Linguistics and Labeling

At 14 months, children accept minimal pairs (words that differ by a single phoneme, like *daw* and *taw*) as labels for the same object. By 20 months, they are less likely to make this mistake due to their experience with phonemes in different word forms. **Thiessen and Yee** (p. 1287) sought to determine how widely 15-month-olds generalize knowledge from previous experience. They find that the effect of training is limited, at least initially, to contexts that are very similar to the initial experience. This suggests that children's early word-form representations are at least partially context specific.

Eyes, Ears, Nose, and Mouth

Decoding facial emotions is critical to allowing us to evaluate others' states of mind and intentions. **Vlamings, Jonkman, and Kemner** (p. 1304) studied 70 Dutch 3- to 8-year-olds to investigate the role played by different brain systems in the perception of faces and facial expressions. Children showed enhanced early visual brain responses to fearful faces compared to neutral faces, as do adults. But for the children, the brain's responses to expression appeared to be driven by a different visual pathway than that used by adults. While adults typically process faces using a visual pathway that relies on coarse information such as shading, distance between eyes, and hairlines (low spatial frequency cues), the children appeared to be using a system that relies on finer details and features (high spatial frequency cues). The finding that children process faces using different pathways than adults has implications for various theories regarding the nature of face processing across development, and may be relevant for our understanding of developmental disorders such as autism.

Family Functioning

Guided by family systems theory, **Sturge-Apple, Davies, and Cummings** (p. 1320) took a longitudinal look at how family relationship patterns affect children's adaptation to school. They studied more than 225 mostly European American families with 6-year-old children over a 3-year period to first identify three types of families—cohesive, disengaged, and enmeshed. Then they determined how these types of families affected children's adjustment to school. Children in disengaged families had the most difficulty adjusting to school, with higher levels of aggression and delinquent behavior in the classroom, and they continued to have these problems over the 3 years of the study. Children in both disengaged and enmeshed families had higher levels of depression and anxiety, and more difficulty emotionally adjusting to school over the course of the study. The research identifies new ways to analyze how families function and offers insights into young children's adjustment to school.

Anne Bridgman

Child Development, July/August 2010, Volume 81, Number 4, Pages 1029–1039

Disasters and Their Impact on Child Development: Introduction to the Special Section

Ann S. Masten
University of Minnesota, Twin Cities

Joy D. Osofsky
Louisiana State University Health Sciences Center, New Orleans

Disasters touch the lives of millions of children every year in many forms. These include natural disasters such as earthquakes, hurricanes, tornadoes, fires, or floods; human-made disasters of armed conflict, genocide, industrial accidents, or terrorism; and disease outbreaks. Interest in the impact of disasters on children among scientists dates back decades to the beginnings of research into risk and resilience in development (Garmezy, 1985; Garmezy & Rutter, 1983). For many years, only a small number of scattered studies of disasters were published on young people, often in the wake of a major catastrophe, such as the Buffalo Creek disaster (Erikson, 1978; Green et al., 1991; Newman, 1976), World War II and the Holocaust (Freud & Burlingham, 1943; Moscovitz, 1985), or a major fire (McFarlane, 1987). In the beginning of the 21st century, with the rise in international terrorism, concerns about flu pandemic, an alarming sequence of natural and human-designed disasters around the world, and globalization of media coverage, there is increasing attention to the consequences of disaster for children and youth (La Greca, Silverman, Vernberg, & Roberts, 2002; Masten & Obradović, 2008; Osofsky, Osofsky, & Harris, 2007; Sagi-Schwarz, Seginer, & Abdeen, 2008). Research into disasters has accelerated despite inherent difficulties in conducting research in the context of catastrophic events. Given the importance of understanding how different types of disasters may impact development for children and families, it was timely to devote a special section of this journal to the impact of disasters on child development.

The goal of the special section on disasters and child development was to provide an opportunity for researchers around the world to examine how disasters of nature and human design might affect children of different ages, experiences, cultures, and contexts, as well as how exposure to a disaster may alter developmental processes or developmental trajectories. Potential contributors were invited to submit a letter of intent by January 15, 2008, and 67 letters for different possible contributions were received. We invited 39 potential contributors to submit full manuscripts for review by May 1, 2008. The 15 articles comprising this special section represent 22% of the submitted letters of intent and 45% of the manuscripts that were submitted for review.

The process of review was complicated by the diversity of the submissions in terms of disaster types and contexts, ages of disaster victims, nationality of authors, disciplinary approach, methods, and focus of the articles (e.g., empirical, conceptual, review, commentary). As editors for this special section, we strove to capture a broad representation of contemporary international scholarship from among the submissions that would reflect the diversity in the field. We were well aware of the challenges of conducting research in the context of disaster (discussed further below) and the constraints posed by ethical and practical dilemmas. In some cases, a study represents rare data obtained by methods that may not meet typical standards of scientific rigor because of the study context (e.g., naturalistic studies in the field or remote locations after a disaster). In all cases, we based the editorial decisions on our judgment of the potential value and unique contribution of the article in the context of current knowledge.

Correspondence concerning this article should be addressed to Ann S. Masten, University of Minnesota, 51 East River Road, Minneapolis, MN, 55455, or Joy D. Osofsky, Louisiana State University Health Sciences Center, 1542 Tulane Avenue, New Orleans, LA 70112. Electronic mail may be sent to amasten@umn.edu or josofs@lsuhsc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0002

The articles in the special section collectively provide a picture of the state of the research into disaster and development in children and adolescents. The diversity of the articles is striking, spanning a broad array of disaster situations, including war and child soldiers, political conflict and terrorism, hurricanes and tsunami, as well as earthquakes and floods. Specific disasters occurring in nine different countries are represented. The articles include many age groups from early childhood through adolescence. Longitudinal studies continue to be rare in this domain of science as are studies that address developmental processes. Most of the articles focus on a behavioral level of function and analysis, although several articles address other levels of analysis, including biological function, relationships, and the recovery context.

In the following sections, we highlight key issues and findings arising from the special section. In the conclusion, we comment on the implications of the special section for future developmental research and applications for improving disaster response and preparedness for the well-being of young people.

Challenges of Research Into Disasters and Child Development

Research in the context of disaster is not for the faint of heart. It is extraordinarily challenging from ethical, conceptual, methodological, and practical perspectives. It can be stressful and sometimes dangerous for the investigators as well as the participants. Special ethical issues must be considered when research is conducted among groups of traumatized survivors. The authors of these articles can be commended for their courage and persistence in the face of such obstacles, as well as their sensitivity and success in executing their work.

Disasters, by their nature, pose great challenges to researchers. They often occur with little advance warning and devastation on a scale that generates many issues for research, ranging from the ethical to the pragmatic. In the immediate aftermath of a catastrophe, survival and basic needs take precedence over research and significant issues arise about the competence of recently traumatized individuals to provide informed consent to research. In addition, research conducted in sites of mass destruction with traumatized people can be both dangerous and traumatizing to those carrying out research. There can be many other kind of dangers, known and unknown, related to ongoing violence or aftershocks, exposure to contagious diseases, lack of supplies, electricity, housing, and many of the modern tools researchers often take for granted, such as offices, computers, mobile telephones, and refrigeration. Migration, chaos, and instability may follow disaster, creating enormous difficulties in tracking individuals for longitudinal studies and often resulting in substantial missing data. These issues, as well as ongoing difficulties in obtaining funding after the immediate impact of the disaster and for follow-up, may contribute to the paucity of longitudinal studies in this area.

Ideally, disaster research would include predisaster assessments, but this standard poses an enormous challenge given the unpredictability of most disasters. In a small number of studies in the literature, investigators have been able to capitalize on research already underway prior to a disaster. In the special section, a study of Kenyan children in a preschool serving 100 children was begun 4 months prior to the eruption of political violence (Kithakye, Morris, Terranova, & Myers, 2010). Follow-up data were obtained 3 months after the onset of violence for 84 of the 92 participants of the preschool study. However, in most cases, the articles of the special section focus on data collected postdisaster, as is typical in most of the extant literature.

Additional design issues confront investigators. Representative samples are difficult to access following disasters and most of the studies focus on samples of convenience or subpopulations. As indicated earlier, samples also may change over time related to instability in the disaster-affected region, leading to missing data. Epidemiological data are rare in this field, for understandable reasons. Therefore, the data provided by Becker-Blease, Turner, and Finkelhor (this issue) offer much-needed information on the prevalence and incidence of disasters in a nationally representative sample of younger children and adolescents for the United States. Data were drawn from the Developmental Victimization Survey conducted in the winter of 2002–2003 to assess the prevalence of various traumatic experiences among American children ages 2–17. About 14% of the young people had experienced some disaster event in their lifetime and around 4% in the past year. Within and across each age group studied (children ages 2–9 by parent report and youth ages 10–17 by self-report), age was associated with higher disaster exposure.

Another challenge is posed by the reality that disasters often unfold in contexts and cultures where research has not been common. One of the most important issues for developmental science in general is the fact that a disproportionate amount of research has been concentrated in more economically advanced societies and contexts convenient to scientists, such as cities and towns, particularly those near academic centers. As a result, investigators trying to study disasters and child development often are faced with challenging issues of measurement and the lack of a database on normative development in many cultures and regions of the world. Additionally, participants in developing countries may not be able to read or have no written language. There may be no relevant instruments in the language or culture of the exposed population, along with a shortage of interpreters, and limited or nonexistent validity data on the value of the tools. This situation also may be confounded by cultural differences in knowledge or support for research as well as differences in the understanding of fundamental concepts of psychological well-being and health. Cultural differences can pose other barriers of access, relevance, and understanding.

These challenges of disaster research have shaped the nature of international research on child development and disasters, and their influence is evident in the special section. It is all too easy to critique the design and methods of disaster research or to underestimate what has been achieved in an investigation if one ignores the reality of these challenges. The special section represents diverse approaches to this array of challenges, with some remarkable examples of success under extraordinarily difficult research conditions. Some investigators worked with local nongovernmental organizations (NGOs) to facilitate data collection. For example, Catani et al. (this issue) partnered with German and Sri-Lankan organizations to train 1,000 teachers in Sri Lanka before they conducted a series of epidemiological surveys of Tamil children before and after the tsunami of December 2004. Also in Sri Lanka, Fernando, Miller, and Berger (this issue) enlisted an NGO partner for their study of disaster-survivors of the tsunami and political violence. Betancourt et al. (this issue) partnered with the International Rescue Committee (IRC) to conduct their ground-breaking longitudinal study of child soldiers of Sierra Leone. This partnership made the study possible; however, the death of the IRC country director midstudy and suspension of their program also brought data collection to a halt before it was completed. In the United States, Vigil, Geary, Granger, and Flinn (this issue) conducted a study of adolescents dislocated by Hurricane Katrina beginning 3 weeks after a government trailer camp opened near Baton Rouge and 2 months after the hurricane. Their study included sampling of saliva from adolescents living in this trailer-park community, as well as a comparison group, providing a rare study of disaster in youth that included biological as well as behavioral levels of assessment. In New Orleans after Katrina, the Osofsky team was able to begin research immediately after school reopened in St. Bernard parish, just 2½ months post-hurricane, because the investigators were local and returned to the disaster area immediately to work on recovery (Kronenberg et al., this issue; Osofsky et al., 2007).

Many of the investigators had to create or translate measures that would be appropriate to the language, culture, and disaster context, relying on bootstrapping methods to demonstrate psychometric reliability or validity. A number of the investigators describe the challenges of translating and back-translating measures, while striving for cultural equivalence, or adapting measures for new contexts (e.g., Betancourt et al., this issue; Catani et al., this issue; Fernando et al., this issue; Kithakye, Morris, Terranova, & Myers, this issue; Klasen et al., this issue; Layne et al., this issue). Betancourt et al. describe the strategies used to create their measures through multiple stages of local consultation, youth focus groups, translation back and forth, and psychometric analyses. Fernando et al. conducted several "comprehensibility checks" with students in pilot schools to refine the survey instruments for their study. Layne et al. enlisted the aid of doctoral students in psychology at the University of Sarajevo to help them review their measures for cultural relevance and language accuracy for their study of Bosnian adolescents who experienced war and political violence following the collapse of the former Yugoslavia.

Most of the studies are cross-sectional rather than longitudinal, with notable exceptions (Betancourt et al., this issue; Kilmer & Gil-Rivas, this issue; Kithakye et al., this issue; Kronenberg et al., this issue). The study by Betancourt et al. represents one of the first longitudinal studies of child soldiers ever reported in the literature. The first prospective longitudinal data on a cohort of male and female child soldiers, this study constitutes a major contribution to a limited literature. Two studies of recovery and resilience among children who experienced Hurricane Katrina by Kronenberg et al. and by Kilmer and Gil-Rivas both include longitudinal data.

The challenges posed for research during the crisis phase of disaster or ongoing conflicts highlight the significance of diverse approaches to important questions for policy, humanitarian assistance, disaster planning, and intervention. These questions include the following: What is the range of human response and how does this vary by development and situation? What can and should be done to help? What works best for whom in which context? What can be harmful? One approach to the conundrum of executing meaningful research under emergency conditions to address such questions is to draw on the expertise of people with extensive field experiences. Ager, Stark, Akesson, and Boothby (this issue) have drawn on the expertise of leaders in humanitarian agencies with extensive disaster and crisis experience to generate data for a "Delphi" consensus analysis on the best practices for care and protection of young people in such crisis situations. Their strategy is more structured and systematic but similar to the convening of consensus panels to identify best practices for diagnosis or intervention (see Hobfoll et al., 2007, for results of a consensus panel on mass trauma intervention). Ager et al. note the congruence of the consensus findings on best practices with the developmental literature on ecosystems of development and resilience.

Highlights of Findings in the Special Section

The conceptual frameworks for the articles in the special section draw heavily on developmental systems theory and many of them focus on concepts of cumulative risk and resilience (Masten & Obradović, 2008). Response to disaster is often linked to the severity of exposure, described as a "dose–response gradient," to previous experiences of trauma or to the conditions of the recovery environment. Congruent with the general risk literature, symptoms or problems often are related to greater cumulative exposure, defined in terms of severity (intensity) or a piling up in time of multiple traumatic experiences. At the same time, striking variability may be observed in the range of behaviors observed in individuals following similar levels of trauma exposure, suggesting that other influences play a role in adaptation to disaster. These include individual differences, as well as differences in the supports or recovery context. Promotive or protective influences that foster resilience hold particular significance for efforts to prepare for disaster and recovery more effectively (Layne et al., 2009; Masten & Obradović, 2008).

Results converge on several important themes for understanding the effects of disaster on development. These include the significance of exposure in terms of nature and dose, the importance of pre- and postdisaster context for understanding disaster response and recovery, protective effects associated with better recovery, and the possible role of age and gender differences both for exposure and response patterns.

Dose Effects, Threshold Effects, and Unpacking Trauma

One of the widely supported findings in the literature on trauma and its consequences is the dose–response effect: As the level of extreme adversity exposure rises or accumulates, there is an increase in symptoms of trauma, behavior problems, mental anguish, and many other kinds of problems observed in children as well as in adults (Norris et al., 2002; Pine, Costello, & Masten, 2005). Similar risk gradients have been observed in diverse studies of cumulative risk and adversity in the developmental literature (Obradović, Shaffer, & Masten, in press; Shonkoff, Boyce, & McEwen, 2009).

Dose–response effects were observed in various forms in the studies of the special issue. Catani et al. (this issue) found that children in Sri Lanka exposed to the 2004 tsunami showed worse adaptation after the disaster in relation not only to more severe exposure but also as a function of higher exposure to other adversities, including the ongoing war, family violence, and other psychosocial risks. Similar findings are reported by Fernando et al. (this issue) in their study of tsunami survivors. Celebi Oncu and Wise (this issue) found differences in stories by children who were directly exposed to the 1999 Kocaeli earthquake 2 years after the disaster compared with children from a distant town in Turkey. Kronenberg et al. (this issue) found that youth with better patterns of mental health 2 and 3 years following Katrina had lower initial trauma exposure and also less recovery adversity. Additionally, the study of political conflict in Kenya, by Kithakye et al. (this issue), indicated that the severity of the disaster experience was associated with more negative and fewer prosocial behaviors.

Dose–response effects were also observed in relation to parent function by Chemtob et al. (this issue) in their study of children following 9/11 in New York City. Preschool children of parents with more symptoms of mental health problems had more problems. These results are consistent with developmental theory suggesting that children who are not protected at the time of disaster by effective caregivers may be particularly vulnerable to disaster effects. One of the mediating pathways by which disasters can harm children is via their effects on parents and parenting quality, particularly among very young children (Masten & Obradović, 2008;

Osofsky, 2004; Osofsky et al., 2007; Pine et al., 2005). Parents have a key role in the protection of children in life-threatening circumstances through actions they take related to preparedness and safeguarding their children, their communications about safety or danger, their instructions or training of children about what to do, and other means, such as their role modeling of adaptive behavior. As a consequence, an important part of disaster preparedness for children involves preparing parents to carry out these roles effectively under very trying circumstances.

It also is important to note that dose–response effects were not always found or sometimes were minimal. The absence of dose–response gradients observed after some time has elapsed since the disaster could reflect recovery processes instead of the absence of dose effects. In their study of adolescents and mothers about 15 months after the 9/11 World Trade Center disaster, Gershoff, Aber, Ware, and Kotler (this issue) found only small effects in one domain (depression) for a dose–response gradient in mental health among the youth in relation to variations in exposure.

There also is some evidence that the dose–response relation may change at very high levels of severe or chronic exposure. For example, results from the study of former child soldiers in Uganda (Klasen et al., this issue) did not show a link between the severity of trauma exposure during their abduction and period of forced service and their postabduction outcomes. It is conceivable that exposure could be so high in some groups that dose no longer is correlated with symptoms because everyone has crossed a threshold that evokes response or overwhelms coping capacity. Beyond this level, differences in vulnerability or protection (moderating influences, such as stress-reactivity or caregiving quality) or differences in the recovery environment may play a role while exposure level may not. Klasen et al. found that posttraumatic outcomes in these child soldiers were associated with lower exposure to adversities after their return, as well as better socioeconomic origins, perceived spiritual support, and less motivation to seek revenge.

The importance of unpacking dose, as well as the nature of the trauma exposure, is underscored by several studies. In their study of young people exposed to war conditions in Bosnia, Layne et al. (this issue) demonstrate that it is feasible and more informative for intervention and theory to disaggregate risk gradients in order to identify differential relations among various risks, promotive factors, or types of trauma exposure and various aspects of outcome. As they note, their results add to a growing body of work focused on differential effects of risks and extreme adversities with an eye toward identifying processes that underlie dose–response patterns and explain how trauma and disaster affect the course of individual functioning and development (Obradović et al., in press). The work of Catani et al. (this issue) highlights the importance of considering the context of the people exposed to disaster, as many of the people of Sri Lanka faced the devastating tidal wave of December 2004, in the context of ongoing war or family violence and economic adversity. Fernando et al. (this issue), in their study of the tsunami survivors of Sri Lanka, argue that it is important to consider the explanatory role of daily stressors as well as disaster exposure.

Betancourt et al. (this issue) describe specifically "toxic" experiences that had lasting and distinctive effects on the child soldiers in this study, including the tragically common experiences of rape and killing others. Young people who perpetrated killing during their time as child soldiers had increasing hostility during the study, whereas those who had experienced rape had more anxiety and hostility, yet also showed more confidence and prosocial attitudes during the course of the longitudinal study. Moreover, they note that female child soldiers who experienced sexual violence face greater stigma than males when they return to the community because of perceptions that these young women are sexually "impure" or "promiscuous." This differential stigma is especially important because Betancourt et al. found that community acceptance was a key protective factor for adjustment of child soldiers after their return to the community.

Results of studies in this issue also emphasize the significance of the adversities in the recovery context or aftermath of disaster. Klasen et al. (this issue), in their study of former child soldiers from Uganda, found that experiences after the children returned were important, with additional exposure to violence (often in the form of domestic or community violence) related to worse outcomes. Kronenberg et al. (this issue) emphasized the role of the postdisaster environment, and particularly the presence of ongoing adversities in the family and community, in their study of students 2–3 years after Hurricane Katrina. The issue of stigma observed for sexually traumatized girls by Betancourt et al. (this issue) represents another form of adversity in the recovery context. However, the significance of the recovery context is not limited to adversity. The special section also underscores the importance of positive features of the recovery context and other protective factors for disaster.

Promotive and Protective Factors

A number of the studies in this section examined variation in the adaptive behavior of the young people studied, often within a resilience framework. Disaster research has played a central role in the history of resilience science, as noted earlier. Studies with a resilience focus typically aim to assess positive as well as negative patterns of adaptation after disaster and also seek to identify the factors or conditions that appear to promote or protect good function during the crisis or recovery period following disaster. Promotive factors predict better outcomes at all levels of risk or adversity (a main effect), whereas protective factors have a greater effect or play a special kind of role when risk or adversity is high (moderating or interaction effect; see Masten, in press).

Moderating effects in response to disaster also can be conceptualized in terms of vulnerability rather than resilience. This distinction can be arbitrary in that it is usually unclear whether the attribute is operating to worsen or improve the outcome, or both. However, when a group of individuals with a particular attribute appear to be particularly susceptible to negative effects of adversity, that attribute is typically described as a vulnerability factor that exacerbates risk. In their commentary in the special section, Peek and Stough (this issue) discuss the vulnerability of children with disabilities in relation to disaster, and concomitantly the vulnerability of communities that do not prepare adequately for meeting the needs of such children during a disaster. Although data are scarce on risk and protective factors for children with disabilities and their families who encounter disasters, children with various disabilities may be particularly endangered by threats encountered in disasters because of specific functional difficulties or sensitivities posed by physical impairments or emotional reactivity related to a disability. They may also be more vulnerable due to lack of planning for children with disabilities during and following disasters. Peek and Stough also note that children with disabilities and their families may be more dependent on special educational or community supports that are disrupted by disaster and therefore more dependent on the community recovery. Yet, it is also conceivable that parents of children with disabilities may develop especially effective strategies and systems for protecting their children in response to the challenges of rearing and protecting a vulnerable child.

Disasters are defined in part by the scope of their impact on the lives of those affected and disasters often threaten the most fundamental adaptive systems for human development (Masten & Obradović, 2008). Children would not be expected to fare well if a tsunami kills their parents and no one is available to care for them, or if they suffer a devastating head injury that impairs subsequent learning and cognition. Extreme psychosocial stress also has the potential to undermine brain development and related cognitive skills, such as memory and executive function through biological stress processes. There is considerable interest in the processes by which psychological or physical adversity experiences become embedded in child development, through a variety of pathways, ranging from the effects of elevated cortisol on the developing brain to the effects of maternal deprivation on attachment, emotional security, mastery motivation, the development of self-regulation, and later relationships (Feder, Nestler, & Charney, 2009; Gunnar & Herrera, in press; Shonkoff et al., 2009).

Several of the articles in the special section address questions on resilience. Generally, the results are consistent with the broader literature on resilience in young people (Luthar, 2006; Masten, 2001, 2007, in press). Kithakye et al. (this issue), in one of the few studies with data on predisaster adjustment, found that self-regulation skills in preschoolers were associated with prosocial behavior in general and had a moderating effect on the relation of exposure severity with prosocial outcomes. The interaction pattern was consistent with other findings in the resilience literature suggesting a protective role of self-regulation skills (Masten, 2007). As indicated previously, community acceptance was important for resilience observed over time in child soldiers by Betancourt et al. (this issue), corroborating other reports on recovery of child soldiers (Boothby, Crawford, & Halperin, 2006). Klasen et al. (this issue) found posttraumatic resilience in former Ugandan child soldiers was associated with fewer guilty cognitions, less desire for revenge, better socioeconomic situation in the family, and greater perceived spiritual support.

In one of the most innovative studies in the special section, Vigil et al. (this issue) examined the potential interacting role of two biological systems—hypothalamic-pituitary-adrenal axis (assessed indirectly by salivary cortisol) and sympathetic nervous system (SNS, assessed indirectly by salivary alpha-amylase)—in relation to function after disaster. Their results are consistent with the possibility that a combination of high cortisol and SNS activity is associated with resilience.

The role of schools in the recovery environment for children has been widely reported in the resilience literature, particularly for war and disaster (Masten & Obradović, 2008), and this perspective is corroborated in the special section. Prompt reestablishment of schooling was one of the most highly endorsed best practices emerging from the consensus study based on the expertise of humanitarian agency leaders with extensive disaster experience (Ager et al., this issue). This theme is also emphasized by Kronenberg et al. (this issue) in the aftermath of Katrina and Betancourt et al. (this issue) in the context of war in Sierra Leone. In the Kronenberg et al. study, schools played a key role in the rebuilding of the community and supporting recovery. In the latter case, staying in school was associated with improved prosocial attitudes and behaviors.

Results also reflect the rising interest in posttraumatic growth in the aftermath of trauma and disaster (see Bonanno, 2004; Kessler, Galea, Jones, & Parker, 2006; Masten & Obradović, 2008; Pat-Horenczyk & Brom, 2007). While the developmental meaning and significance of "better" function during or after disaster remains uncertain, there is growing attention to this kind of phenomenon among young people and adults. In most situations, pretrauma functioning is unknown; therefore, it is unclear how to distinguish normal growth and development from posttrauma growth, and the meaning of better functioning and its long-term possible cost for a child is unknown. The study by Kilmer and Gil-Rivas (this issue) addressed the complex questions of posttraumatic growth in children. While the findings from the Kilmer and Gil-Rivas study are preliminary (and, typical of most disaster studies, did not include predisaster data), these authors suggest the potential benefit of working with children to support relevant competencies and cognitive processes that may foster posttraumatic growth. Their observation that rumination and distressing thoughts were associated with posttraumatic growth raises the intriguing possibility that discomfort in combination with reflective processes could lead to positive changes in the aftermath of trauma.

Gender and Age Differences

The literature on disaster exposure and response in children and adolescents presents a complex picture of gender and age effects, and the studies of this special section do as well. Numerous methodological issues arise in the interpretation of both gender and age effects. Parents and teachers typically are the informants for young children, whereas adolescents often self-report on their own symptoms or well-being, which confounds response with the respondent and most likely gender of the informant with gender of target as women are more likely than men to be the informants for young children. Mothers, for example, may report more of some symptoms in children than objective observers might report and different symptoms than a child himself or herself was able to report. Similarly, when female adolescents report more symptoms than adolescent males, it is not clear whether they experience more symptoms or simply are more willing to admit them.

Following disasters, it is often difficult to find a comparable comparison group, limited measures are available with norm-referenced data on the affected population, and there is rarely any predisaster assessment on the study sample. This situation makes it difficult to know if reported gender or age differences observed among the disaster-exposed group differ from what one might have found in nonexposed young people.

Disaster and trauma studies often find that females report more symptoms (Tolin & Foa, 2006). However, female adolescents and adults generally report or disclose more symptoms of distress, regardless of whether they have experienced trauma (Crick & Zahn-Waxler, 2003). In the Kronenberg et al. (this issue) study following Katrina, female students (4th through 12th grade) reported more depression and trauma symptoms than males, although these investigators noted that it is difficult to know whether these gender differences reflect differential reporting or experiencing of these symptoms. Moreover, without a normative or nonexposed reference group, it is difficult to discern if initial post-Katrina scores reflect elevated symptom levels, although the subsequent decline suggested that this might have been the case. The study by Vigil et al. (this issue) was one of the few to compare a disaster-exposed group (adolescents who experienced Katrina) to a comparable group of nonexposed young people (of comparable sociodemographic background). Results were complex in that Katrina-exposed females appeared to report particularly high symptoms of depression while Katrina-exposed males reported particularly low levels of aggression.

In war and disaster, it also is difficult to sort out the meaning of gender differences even when they are found, because the experiences of males and females may differ in significant ways and the cultural meaning of the experiences may also differ by gender (see Bal, 2008; Barber, 2008; Norris, Perilla, Ibañez, & Murphy, 2001; Pfefferbaum, 1997). As noted earlier, Betancourt et al. (this issue) found that female child soldiers reported more rape experiences than the male child soldiers and that sexual violence held greater stigma for females in the recovery community context.

In the disaster literature, differential vulnerabilities and protective factors have been noted in relation to age for some time (Masten, Best, & Garmezy, 1990). Younger children have been viewed as more vulnerable to the loss or declines in quality of caregiving because of their dependence on care. Young children, on the other hand, are protected in some ways by their lack of exposure to and understanding of disaster as it unfolds, especially when their environment remains relatively stable with support from their families (Osofsky, 2004). Adolescents are viewed as at greater risk of high exposure because of their place in the world, their friends who may have been exposed, and their greater awareness of what is happening and its future significance. Older children and adolescents also are at higher risk of certain dangers (e.g., forced labor, rape, military service). Conversely, older youth have better problem-solving skills, more coping capacity, more social support outside the family, and more capacity to survive on their own than younger children. Thus, the vulnerability and capacity for resilience change over the course of development and also vary with the circumstances of the individual in a disaster.

Age differences in exposure, experiences during and after disaster, and adjustment following disaster were found in several studies in this special section, although the findings are not consistent. Exposure to lifetime disaster was generally higher among older children in the epidemiological survey by Becker-Blease et al. (this issue). Klasen et al. (this issue) observed more trauma and symptoms among the older of the former child soldiers in Uganda, whereas Betancourt et al. (this issue) did not find comparable effects in their study of child soldiers from the conflict in Sierra Leone. Age was unrelated in the latter study to most outcomes, except that older returning soldiers in the Betancourt et al. study experienced less family acceptance than younger children, perhaps because older children evoked less sympathy or more fear. Kronenberg et al. (this issue) found that the younger students (ages 9–11) reported more depression and posttraumatic stress disorder symptoms.

Given the interplay among exposure, development, and individual differences, it is a complex matter to sort out the meaning of observed age-related differences, as noted by Layne et al. (this issue) in their discussion. Indeed, in many respects, the empirical evidence on disaster in relation to development is sparse. Despite numerous signs of progress on multiple fronts in the research on children in disasters, nicely reflected in this collection of articles, developmentally informative data remain limited.

Conclusions

Implications for Research

The special section highlights the challenges, progress, and gaps in research on disaster and child development. Formidable obstacles have impaired progress. These range from the inherent challenges of conducting research under disaster-affected conditions to obtaining rapid funding, especially for those researchers who reside in the disaster-affected areas. However, there is also a profound shortage of suitable standardized and culturally appropriate measures in many of the regions where disasters occur, reflecting a much broader issue in developmental science. There are many more studies and validated measures for developed than developing regions and dominant cultural groups compared with minority groups (Quintana et al., 2006). Building a broader science of child development across cultures and regions would be extremely helpful to research in many domains, and it is critically important for disaster research.

Longitudinal data are scarce, particularly with regard to studies with predisaster baselines, despite the need for information on what may be most helpful to whom and when after disaster. Greater attention may need to be directed toward building national and international supports and structures for disaster research, including technical support and collaborations among teams of developmental scientists, humanitarian service providers, and local leaders or disaster responders.

There continues to be very little attention to developmental issues in disaster research beyond the most basic role of age. Research is needed on the role of developmental timing for exposure, vulnerabilities and protective influences, and patterns of response over the near term and long term. In addition, there is a great need for more developmentally informed and informative research into the effectiveness of prevention and recovery strategies for disaster-affected children and adolescents, especially with regard to considerations of development, gender and cultural differences, and the nature of the disaster exposure. Layne et al. (this issue) provide a compelling argument for research that attempts to unpack the complex interplay among development, trauma exposure, resources, and protective or vulnerability processes, particularly in order to inform intervention. They assert that a more differentiated understanding of ''who is at risk for what outcomes via which pathways of influence'' (p. 1072) will inform evidence-based clinical decisions. Developmental considerations need to be at the center of the unpacking mission.

Implications for Disaster Preparation and Response

The existing literature on disasters and child development and the broad array of articles in this special section provide important information not only for developmental science but also for efforts to address the needs of children and youth in the context of disaster. While there are many gaps in the research, disasters continue to unfold worldwide, and it is reasonable to consider the implications of current findings from the special section and earlier work for efforts to help the many young victims of disaster. There is some consistency in the findings that may afford helpful guidance in preparing for disaster and promoting resilience in its wake.

By definition, it is not possible to be fully ''prepared'' for a disaster. Nonetheless, the findings to date, including data from articles in this section, suggest that communities can plan and prepare to support the mobilization of family and community resources to protect children and promote resilience in situations of overwhelming adversity. The lessons for preventive intervention from the growing body of research on disasters and children, as highlighted in the special section, can be summarized in terms of developmental guidelines as follows (see also Hobfoll et al., 2007; Masten & Obradović, 2008; Norris, Steven, Pfefferbaum, Wyche, & Pfefferbaum, 2008; Norris & Stevens, 2007; Osofsky et al., 2007; Pine et al., 2005).

1. Protect and restore the secure base of attachment relationships as soon as feasible. Prepare for evacuations and relocations that include keeping children together with their caregivers. Reunite children and adults who have attachment bonds if they have to be separated. Provide nurturing caregivers for children who have lost their caregivers.
2. Train first responders in the range of developmental responses to trauma that can be expected for children of different ages.
3. Remember that first responders for children include parents, teachers, and day-care providers in addition to emergency responders.
4. Support normalizing routines, activities, and contexts for children after disasters, including opportunities for them to play and learn, and the restoration of functional schools and other community organizations that serve children and their families.
5. Attend to community resilience and the function of cultural and community practices that support families and their children, including support for spiritual practices.
6. Support and restore meaningful relationships and opportunities to be effective in play, school, work, or recovery activities. Such activities may nurture hope, meaning, and a sense of agency, both in adults who care for children and among older children and adolescents themselves.

Perhaps the most important role for developmental scientists in disaster response, however, is for them to engage more fully in the process of building and applying knowledge. Scientists with developmental knowledge rarely are included in decision making related to recovery and rebuilding following disasters, while scientists, in turn, may not reach out to help as effectively as possible. In the United States, the National Child Traumatic Stress Network (Pynoos et al., 2008) can play an important role in brokering this bidirectional process for disaster planning and response. Internationally, Ager et al. (this issue) have articulated the requirements for developmental scientists to contribute to the humanitarian efforts on behalf of children in disasters. These include research engagement in more diverse cultural settings around the world, including those in crisis, and more effective translational efforts to communicate their findings to practitioners and policymakers.

The work of the scientists who have contributed to this special section on disasters and child development illustrates what can be done under extraordinarily challenging conditions. At the same time, the special section also underscores the limits of present knowledge and potential for further research contributions by developmental scientists in this area. Developmental scholars have a vital role to play in shaping future science and its applications to reduce the negative effects of disaster on human development and promote adaptive responses and recovery in young people, their families, communities, and societies.

References

Ager, A., Stark, L., Akesson, B., & Boothby, N. (2010). Defining best practice in care and protection of children in crisis-affected settings: A Delphi study. *Child Development, 81*, 1271–1286.

Bal, A. (2008). Post-traumatic stress disorder in Turkish child and adolescent survivors three years after the Marmara earthquake. *Child and Adolescent Mental Health, 13*, 134–139.

Barber, B. K. (2008). Contrasting portraits of war: Youths' varied experiences with political violence in Bosnia and Palestine. *International Journal of Behavioral development, 32*, 298–309.

Becker-Blease, K. A., Turner, H. A., & Finkelhor, D. (2010). Disasters, victimization and children's mental health. *Child Development, 81*, 1040–1052.

Betancourt, T. S., Borisova, I. I., Williams, T. P., Brennan, R. T., Whitfield, T. H., de la Soudiere, M., et al. (2010). Sierra Leone's former child soldiers: A follow-up study of psychosocial adjustment and community reintegration. *Child Development, 81*, 1077–1095.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we under-estimated the human capacity to thrive after extremely aversive events? *American Psychologist, 59*, 20–28.

Boothby, N., Crawford, J., & Halperin, J. (2006). Mozambique child soldier life outcome study: Lessons learned in rehabilitation and reintegration efforts. *Global Public Health, 1*, 87–107.

Catani, C., Gewirtz, A. H., Wieling, E., Schauer, E., Elbert, T., & Neuner, F. (2010). Tsunami, war, and cumulative risk in the lives of Sri Lankan school children. *Child Development, 81*, 1176–1191.

Celebi Oncu, E. C., & Wise, A. M. (2010). The effects of the 1999 Turkish earthquake on young children: Analyzing traumatized children's completion of short stories. *Child Development, 81*, 1161–1175.

Chemtob, C. M., Nomura, Y., Rajendran, K., Yehuda, R., Schwartz, D., & Abramovitz, R. (2010). Impact of maternal posttraumatic stress disorder and depression following exposure to the September 11 attacks on preschool children's behavior. *Child Development, 81*, 1129–1141.

Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: Current progress and future challenges. *Developmental Psychopathology, 15*, 719–742.

Erikson, K. (1978). *Everything in its path: Destruction of community in the Buffalo Creek Flood*. New York: Simon & Schuster.

Feder, A., Nestler, E. J., & Charney, D. S. (2009). Psychobiology and molecular genetics of resilience. *Nature Reviews Neuroscience, 10*, 446–457.

Fernando, G. A., Miller, K. E., & Berger, D. E. (2010). Growing paints: The impact of disaster-related and daily stressors on the psychological and psychosocial functioning of youth in Sri Lanka. *Child Development, 81*, 1192–1210.

Freud, A., & Burlingham, D. T. (1943). *War and children*. New York: Medical War Books.

Garmezy, N. (1985). Stress-resistant children: The search for protective factors. In J. E. Stevenson (Ed.), *Recent research in developmental psychopathology. Journal of Child Psychology and Psychiatry Book* (Suppl. No. 4, pp. 213–233). Oxford, UK: Pergamon Press.

Garmezy, N., & Rutter, M. (1983). *Stress, coping, and development*. New York: McGraw-Hill.

Gershoff, E., Aber, J. L., Ware, A., & Kotler, J. (2010). Exposure to 9/11 among youth and their mothers in New York City: Enduring associations with mental health and sociopolitical attitudes. *Child Development, 81*, 1142–1160.

Green, B. L., Korol, M., Grace, M. C., Vary, M. G., Leonard, A. C., Gleser, G. C., et al. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 945–951.

Gunnar, M. R., & Herrera, A. M. (in press). The neurobiology of stress and development. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology*. New York: Oxford University Press.

Hobfoll, S. E., Watson, P., Bell, C. C., Bryant, R. A., Brymer, M. J., Friedman, M. J., et al. (2007). Five essential elements of immediate and mid-term mass trauma intervention: Empirical evidence. *Psychiatry, 70*, 283–315.

Kessler, R. C., Galea, S., Jones, R. T., & Parker, H. A. (2006). Mental illness and suicidality after Hurricane Katrina. *Bulletin of the World Health Organization, 84*, 930–939.

Kilmer, R., & Gil-Rivas, V. (2010). Exploring posttraumatic growth in children impacted by Hurricane Katrina: Correlates of the phenomenon and developmental considerations. *Child Development, 81*, 1211–1227.

Kithakye, M., Morris, A. S., Terranova, A. M., & Myers, S. S. (2010). The Kenyan political conflict and children's adjustment. *Child Development, 81*, 1114–1128.

Klasen, F., Oettingen, G., Daniels, J., Post, M., Hoyer, C., & Adam, H. (2010). Posttraumatic resilience in former Ugandan child soldiers. *Child Development, 81*, 1096–1113.

Kronenberg, M. E., Hansel, T. C., Brennan, A. M., Lawrason, B., Osofsky, H. J., & Osofsky, J. D. (2010). Children of Katrina: Lessons learned about post-disaster symptoms and recovery patterns. *Child Development, 81*, 1241–1259.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (Eds.). (2002). *Helping children cope with* *disasters and terrorism*. Washington, DC: American Psychological Association.

Layne, C. M., Beck, C. J., Rimmasch, H., Southwick, J. S., Moreno, M. A., & Hobfoll, S. E. (2009). Promoting "resilient" posttraumatic adjustment in childhood and beyond: "Unpacking" life events, adjustment trajectories, resources, and interventions. In D. Brom, R. Pat-Horenczyk, & J. Ford (Eds.), *Treating traumatized children: Risk, resilience, and recovery* (pp. 13–47). New York: Routledge.

Layne, C. M., Olsen, J. A., Baker, A., Legerski, J.-P., Isakson, B., Pašalić, A., et al. (2010). Unpacking trauma exposure risk factors and differential pathways of influence: Predicting post-war mental distress in Bosnian adolescents. *Child Development, 81*, 1053–1076.

Luthar, S. S. (2006). Resilience in development: A synthesis of research across five decades. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology (2nd ed.): Vol. 3 Risk, Disorder, and Adaptation* (pp. 739–795). Hoboken, NJ: Wiley and Sons.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*, 227–238.

Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19*, 921–930.

Masten, A. S. (in press). Risk and resilience in development. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology*. New York: Oxford University Press.

Masten, A. S., Best, K. M., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*, 425–444.

Masten, A. S., & Obradović, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society, 13*(1), 9 [Online] URL: http://www.ecologyandsociety.org/vol13/iss1/art9/.

McFarlane, A. C. (1987). Posttraumatic phenomena in a longitudinal study of children following a natural disaster. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 764–769.

Moscovitz, S. (1985). Longitudinal follow-up of child survivors of the Holocaust. *Journal of the American Academy of Child Psychiatry, 24*, 401–407.

Newman, C. J. (1976). Children of disaster: Clinical observations at Buffalo Creek. *The American Journal of Psychiatry, 133*, 306–312.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. (2002). 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry, 65*, 207–239.

Norris, F. H., Perilla, J. L., Ibañez, G. E., & Murphy, A. D. (2001). Sex differences in symptoms of posttraumatic stress: Does culture play a role? *Journal of Traumatic Stress, 14*, 7–28.

Norris, F. H., Steven, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology, 41*, 127–150.

Norris, F. H., & Stevens, S. P. (2007). Community resilience and the principles of mass trauma intervention. *Psychiatry, 70*, 320–328.

Obradović, J., Shaffer, A., & Masten, A. S. (in press). Risk in developmental psychopathology: Progress and future directions. In L. C. Mayes & M. Lewis (Eds.), *The environment of human development: A handbook of theory and measurement*. New York: Cambridge University Press.

Osofsky, J. D. (Ed.). (2004). *Young children and trauma: Interventions and treatment*. New York: Guilford.

Osofsky, J. D., Osofsky, H. J., & Harris, W. W. (2007). Katrina's children: Social policy considerations for children in disasters. *Social Policy Report, 21*(1), 3–18.

Pat-Horenczyk, R., & Brom, D. (2007). The multiple faces of post-traumatic growth. *Applied Psychology, 56*, 379–385.

Peek, L., & Stough, L. M. (2010). Children with disabilities in the context of disaster: A social vulnerability perspective. *Child Development, 81*, 1260–1270.

Pfefferbaum, B. (1997). Posttraumatic stress disorder in children: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1503–1511.

Pine, D. S., Costello, J., & Masten, A. S. (2005). Trauma, proximity, and developmental psychopathology: The effects of war and terrorism on children. *Neuropsychopharmacology, 30*, 1781–1792.

Pynoos, R. S., Fairbank, J. A., Steinberg, A. M., Amaya-Jackson, L., Gerrity, E., & Mount, M. L. (2008). The National Child Traumatic Stress Network: Collaborating to improve the standard of care. *Professional Psychology: Research and Practice, 39*, 389–395.

Quintana, S. M., Aboud, F. E., Chao, R. K., Contreras-Grau, J. C., Cross, W. E., Hudley, C., et al. (2006). Race, ethnicity, and culture in child development: Contemporary research and future directions. *Child Development, 77*, 1129–1141.

Sagi-Schwarz, A., Seginer, R., & Abdeen, Z. (2008). Chronic exposure to catastrophic war experiences and political violence: Links to the well-being of children and their families. Introduction to the Special Issue. *International Journal of Behavioral Development, 32*, 257–259.

Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities. *Journal of the American Medical Association, 301*(21), 2252–2259.

Tolin, D. F., & Foa, E. B. (2006). Sex Differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin, 132*, 959–992.

Vigil, J. M., Geary, D. C., Granger, D. A., & Flinn, M. V. (2010). Sex differences in salivary cortisol, alpha-amylase, and psychological functioning following Hurricane Katrina. *Child Development, 81*, 1228–1240.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1040–1052

Disasters, Victimization, and Children's Mental Health

Kathryn A. Becker-Blease
Oregon State University

Heather A. Turner and
David Finkelhor
University of New Hampshire

In a representative sample of 2,030 U.S. children aged 2–17, 13.9% report lifetime exposure to disaster, and 4.1% report experiencing a disaster in the past year. Disaster exposure was associated with some forms of victimization and adversity. Victimization was associated with depression among 2- to 9-year-old disaster survivors, and with depression and aggression among 10- to 17-year-old disaster survivors. Children exposed to either victimization only or both disaster and victimization had worse mental health compared to those who experienced neither. More research into the prevalence and effects of disasters and other stressful events among children is needed to better understand the interactive risks for and effects of multiple forms of trauma.

Disasters are relatively common events that disrupt children's basic needs including access to food, water, shelter, and primary caregivers. Much of what we know about disasters and children comes from samples of children exposed to a relatively few, large-scale disasters rather than from representative samples across the population (e.g., Centers for Disease Control, 2005; Landrigan et al., 2008; see Norris et al., 2002, for a review). In addition, we know little about how exposure to disasters is associated with exposure to other kinds of stressful events. In this study, we investigate the prevalence and effects of disasters and other stressful events in a representative sample of U.S. children.

Definitions

In this article, we use the term *disaster* to refer to one-time or ongoing events of human or natural cause that lead groups of people to experience stressors including the threat of death, bereavement, disrupted social support systems, and insecurity of basic human needs such as food, water, housing, and access to close family members. We use the term *victimization* to refer to "harms that occur to individuals because other *human actors* behaving in ways that *violate social norms*" (italics in original, Finkelhor & Kendall-Tackett, 1997, p. 2). Victimization includes maltreatment, sexual and physical abuse, neglect, exposure to domestic violence, and other crimes such as theft and peer and sibling physical assault. Finally, we use the term *adversity* to refer to other (nonvictimization, nondisaster) events that have been associated with poorer mental health, including having an incarcerated or substance-abusing parent. Disasters, victimization, and other adverse events are interrelated in complex ways. Exposure to any one type of event could put children at risk to experience other types of events, or affect how children respond if and when they experience another kind of event. Additional environmental and genetic factors could be implicated both in risk of exposure and for outcomes following exposure to any of these events.

Disasters

General population studies indicate that many people in the United States experience disasters each year. Adults self-report a lifetime prevalence rate of 11%–30% (Briere & Elliott, 2000; Goldberg &

For the purposes of compliance with Section 507 of PL 104-208 (the "Stevens Amendment"), readers are advised that 100% of the funds for this program are derived from federal sources (this project was supported by Grants 1999-JP-FX-1101 and 2002-JW-BX-0002 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice). The total amount of federal funding involved is $584,549. Points of view or opinions in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice. The authors wish to thank Richard Ormrod for help with data management, and John Boyle and Patricia Vanderwolf for help in data collection.

Correspondence concerning this article should be addressed to Kathryn A. Becker-Blease, Department of Psychology, Oregon State University, Corvallis, OR 97331. Electronic mail may be sent to kathryn.blease@oregonstate.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0003

Freyd, 2006; Green & Solomon, 1995; Norris et al., 2002) depending in part whether disaster is defined exclusively as natural disasters, or also includes human-caused incidents. To the best of our knowledge, this study is the first to obtain a national prevalence estimate for disaster exposure among children in the United States.

Numerous researchers have examined children's reactions to large-scale disasters worldwide. From these studies, we know that disasters are implicated in a range of physical effects, including traumatic brain injury, low birth weight, and respiratory, musculoskeletal, and other physical symptoms (Dirkzwager, Kerssens, & Yzermans, 2006; Keenan, Marshall, Nocera, & Runyan, 2004; Landrigan et al., 2008; Rath et al., 2007). Psychological sequalae include posttraumatic stress symptoms, depression, anxiety, behavior problems, sleep problems, and learning problems (Norris et al., 2002; Stein et al., 2004; Weems et al., 2007). Based on a review of many studies, Norris et al. (2002) concluded that school-age children are more likely than adults to experience severe impairment following disasters.

A few studies point to differential effects of disasters for adolescents compared to younger children. Increased aggression and enuresis among 2- to 9-year-olds have been reported 6 months after a flood (Durkin, Khan, Davidson, Zaman, & Stein, 1993) and preschoolers demonstrate increased fear of storms, concern for others, and other signs of disaster experience in their play up to 1 year following a hurricane (Saylor, Swenson, & Powell, 1992). Other studies of children who were under age 5 at the time of exposure to nuclear accidents show few long-term effects (Bromet et al., 2000; Cornely & Bromet, 1986). Dirkzwager et al. (2006) reported that children aged 4–12 reported increased sleep problems relative to controls, whereas adolescents aged 13–18 reported increased anxiety symptoms compared to controls. Green et al. (1991) studied children who experienced a dam collapse 2 years prior. They report that children aged 2–8 reported fewer posttraumatic stress disorder symptoms than children aged 9–15. Most symptoms had disappeared 17 years after the disaster, but some new symptoms (suicidal ideation and substance abuse) arose as participants aged (Green et al., 1994). Based on the available evidence, young children have been found to show less severe symptoms of shorter duration than older children and adolescents, but not enough data are available to speak definitively about age effects among children, particularly for those under 5.

Victimization

Studies investigating the effects of various forms of victimization (Finkelhor, Ormrod, & Turner, 2007) and adverse childhood experiences (Dong et al., 2004) have demonstrated that many people who experience one type of stressful event are likely to experience other types. When multiple stressful events are not considered simultaneously, consequences of multiple events may be attributed to the single event that is the focus of study. There is a particular risk of this effect with disaster research because most disaster research seeks samples after a particular event for the specific purpose of determining the effects of that one particular event. The data set used in this study allows us to estimate the effects of a wide range of stressful events, including disasters.

Disasters, Victimization, and Other Adversity

It is easy to imagine that risk for some stressful events puts children at risk for similar types of events. Poverty, parenting capacity, social support, and many other environmental factors may lead to the empirical evidence demonstrating associations among types of maltreatment and victimization. Why would we predict an association between exposure to victimization and disasters? At least two mechanisms could result in an association between disasters and other stressful events. First, the stress of the disaster may impair parents' ability to care for their children. As reviewed by Norris et al. (2002), adults exposed to disaster may experience sleep disturbance, a worsening of mental health symptoms, reduced social support, and increased use of alcohol following disasters. Further, several studies have pointed to increases in child abuse following natural disasters, although this finding is not universal (Curtis, Miller, & Berry, 2000; Keenan et al., 2004).

It is also possible that some of the same risk factors are associated with both disasters and victimization. In the geography and community development literature, vulnerability to disaster is conceptualized as the interaction between risk of disaster exposure and the ability to prepare for and respond to disasters (Cutter & Finch, 2008; Masozeraa, Bailey, & Kerchner, 2007). For the purposes of this article, we are mainly concerned with risk of exposure, but the overlap of concepts is relevant to our definition of *disaster*, as well as our survey respondents' definitions. People who experience an ice storm for which both they and their community are prepared are less likely to self-report disaster exposure than another person who survived a storm that resulted in injuries and death. In another example, a homeowner living in a well-maintained home is at less risk for experiencing a major fire than is someone living in an apartment complex with inadequate fire alarms and escape routes. For these reasons and others, socioeconomic status (SES), urbanity, and race and ethnicity are persistent predictors of vulnerability to disaster (Cutter & Finch, 2008). These same factors are consistent predictors of exposure to victimization (Finkelhor et al., 2007; Stein, Jaycox, Kataoka, Rhodes, & Vestal, 2003). Likewise, some of the same factors that underlie risk of disaster exposure, also put children at risk for poorer mental health outcomes. Parents who have more resources are not only more able to locate safe shelter out of the way of a natural disaster but also in a better position to help children cope with the event and obtain professional mental health services if required.

Research Questions and Hypotheses

1. How many children are affected by disasters in the United States?
2. Is risk for disaster exposure associated with risk for exposure to other stressful events? We hypothesized that risk for disaster would be associated with risk of physical abuse, exposure to domestic and community violence, and other adversity. We made no hypotheses about the causality of this association.
3. To what extent is victimization history associated with mental health symptoms among disaster victims? We hypothesized that among disaster victims, victimization would be a significant predictor of mental health.
4. How do mental health symptoms compare for children who have experienced disasters, victimization, both disaster and victimization events, and neither disaster nor victimization events? We hypothesized that those who had experienced neither a disaster nor victimization would report the least mental health symptoms, those who had experienced both would report the most symptoms, and those who experienced either a disaster or victimization would have intermediate scores.

Method

Participants

This research is based on data from the Developmental Victimization Survey (DVS), designed to obtain 1-year incidence estimates of a comprehensive range of childhood victimizations across gender, race, and developmental stage. The survey, conducted between December 2002 and February 2003, assessed the experiences of a nationally representative sample of 2,030 children aged 2–17 living in the contiguous United States. The survey was conducted in English, and 1% of those contacted were unable to complete the survey due to a language barrier. The interviews with parents and youth were conducted over the phone by the employees of an experienced survey research firm specially trained to talk with children and parents. Telephone interviewing is a cost-effective methodology (Weeks, Kulka, Lessler, & Whitmore, 1983) that has been demonstrated to be comparable in reliability and validity with in-person interviews, even for sensitive topics (Bajos, Spira, Ducot, & Messiah, 1992; Bermack, 1989; Czaja, 1987; Marin & Marin, 1989). The methodology is also used to interview youth in the U.S. Department of Justice's National Crime Victimization Survey (Bureau of Justice Statistics, various years) and in a variety of other epidemiological studies of youth concerning violence exposure (Hausman, Spivak, Prothrow-Stith, & Roeber, 1992).

The sample selection procedures were based on a list-assisted random digit dial telephone survey design. This design increases the rate of contacting households with children aged 2–17 while decreasing the rate of dialing business and nonworking numbers. Experimental studies have found this design to decrease standard errors relative to the standard Mitofsky–Waksberg (Waksberg, 1978) method while producing samples with similar demographic profiles (Brick, Waksberg, Kulp, & Starer, 1995; Lund & Wright, 1994).

A short interview was conducted with an adult caregiver (usually a parent) to obtain family demographic information. Because the questions in the survey are of a sensitive nature, interviewers explained the study honestly and carefully up front. Care was taken to not refer to the study as a "crime" study, as many victims of crimes do not label their experiences as a *crime*. Instead, children's "safety" was emphasized.

Specifically, parents of 2- to 9-year-olds were told:

> This study is being conducted by the [Name of University] to better understand problems facing children today. For this study, we will be interviewing 2000 families across the country to find out about stressful events that happen to some children, and how schools and various agencies may better protect kids from dangerous situations. Your family was selected at random to represent families with children aged 2–17. Your interview will take about 30 minutes. We will be asking you about things that may have happened in your child's school, neighborhood, or home, and about how your child's health has been lately. Some of the questions involve sensitive issues, such as whether your child has ever experienced violence and whether your child has ever experienced unwanted sexual advances.

Parents of 10- to 17-year-olds were told:

> In learning about child safety for children, we would like to get the input of both children and their parents. It is particularly important for us to find out what kinds of situations *kids* consider dangerous, whether they have ever encountered these situations, and what they know about how to avoid or handle these situations . . . We will be asking him/her about things that may have happened in your child's school, neighborhood, or home, and about how your child's health has been lately. . .
>
> Some of the questions involve sensitive issues, such as whether your child has ever experienced violence and whether your child has experienced unwanted sexual advances.

Before beginning the survey, 10- to 17-year-olds were told:

> Your age group (x-year-olds) is part of national sample of 2,000 boys and girls selected completely at random. You will be representing all kids in the United States in your age group. This is an important study because you'll be able to tell us what kinds of dangerous situations kids face these days. What you and the other kids tell us will help us keep kids safe. Your interview will take about 20–30 minutes. We will be asking you about things that may have happened in your school, neighborhood, or home. We'll also ask about how your health has been lately. Some of the questions involve sensitive issues, such as whether you have experienced violence.

One child was randomly selected from all eligible children living in a household by selecting the child with the most recent birthday. If the selected child was 10–17 years old, the main telephone interview was conducted with the child. If the selected child was 2–9 years old, the interview was conducted with the caregiver who "is most familiar with the child's daily routine and experiences." Caregivers were interviewed as proxies for this age group because the ability of children under the age of 10 to be recruited and participate in phone interviews of this nature has not been well established (Hausman et al., 1992; Waksberg, 1978), yet such children are still at an age when parents tend to be well informed about their experiences both at and away from home. In 68% of these caretaker interviews, the caretaker was the biological mother, in 24% the biological father, and in 8% some other relative or caretaker.

Data were collected using a computer-assisted telephone interview system (CATI). The use of CATI minimizes recording errors and provides substantial quality control benefits. For this survey, only interviewers who had extensive experience interviewing children and in addressing sensitive topics were chosen. Interviewers then went through extensive training on the questionnaire and interview protocol.

Up to 13 callbacks were made to select and contact a respondent and up to 25 callbacks were made to complete the interview. Consent was obtained prior to the interview. In the case of a child interview, consent was obtained from both the parent and the child. Respondents were promised complete confidentiality and were paid $10 for their participation. Telephone interviewers screened 7,907 households. Of those, 5,011 households did not have an eligible child, 511 declined to participate, and 350 agreed to participate at another time but were unable to be reached by the end of the study. Of the 2,896 eligible households screened, interviews were completed with 2,035 participants, representing 70.3% of eligible households. The final sample consisted of 1,000 children aged 10–17 and 1,030 caregivers of children aged 2–9.

All procedures were authorized by the Institutional Review Board of the University of New Hampshire. Both universal and targeted interventions were used to assist children who were potentially at risk for abuse, or harm to self or other. All participants were given the phone number to the Girls and Boys Town National Hotline. The U.S. Department of Health and Human Services describes hotline this way:

> The Girls and Boys Town National Hotline is a 24-hour crisis, resource and referral line. Accredited by the American Association of Suicidology, the Hotline has been in operation since 1989 and is staffed by trained counselors who can respond to questions every day of the week, 365 days a year. Approximately 250,000 callers contact the hotline annually to receive help relating to being suicidal, physically or sexually abused, on the run, addicted, threatened by gang violence, fighting with a parent, a parent frustrated by a child, scared of a spouse, or faced with an overwhelming challenge. (Healthfinder.gov, 2008)

Additionally, the CATI system was programmed to flag automatically cases in which a child could have been in danger. Interviewers also flagged cases that presented a concern and provided brief notes that could be used to better understand the situation. Children or parents who disclosed a situation of serious threat or ongoing victimization were recontacted by a clinical member of the research team, trained in telephone crisis counseling, whose responsibility was to stay in contact with the respondent until the situation was resolved or brought to the attention of appropriate authorities. The kinds of concerns varied widely and included depressed children who were seeking treatment, teenagers who were slapped once or twice by a parent, abused children who no longer had contact with the abuser, and adolescents who reported ''pressure'' to have sex from a romantic partner. In no case did it become necessary for the phone counselor to report suspected abuse of a child.

Measures

Demographic variables. All demographic information was obtained from the parent during the initial parent interview. Parents provided the child's age (in years), race and ethnicity (coded into four groups: White non-Hispanic, Black non-Hispanic, Hispanic any race, or Other Race non-Hispanic) and current family structure (coded into three groups: two biological or adoptive parents, one biological and a married or unmarried partner, or single-parent family). Participants also reported 2002 household income from wages, public assistance and child support (10 categories: $5000 or less through more than $100,000), and the highest education completed (for the parent with the most education, 11 categories: grade school through graduate degree).

Lifetime victimization. Victimization data were collected using the Juvenile Victimization Questionnaire (JVQ; Hamby, Finkelhor, Ormrod, & Turner, 2004). The JVQ was designed to assess a wide range of childhood victimization experiences, across childhood and adolescence. The same questions were used for self-report interviews with 10- to 17-year-olds and parent-report interviews for 2- to 10-year-olds, except for slight wording changes as needed (e.g., ''your child'' substituted for ''you'').

Concerns about the validity of self- and parent-reports take two forms. Some are concerned that respondents may underreport victimization events to avoid culpability, due to stigma, or due to forgetting. Others are concerned that respondents might overreport victimization events, to explain problems in their lives or to assist with child custody claims. It is possible that these motives were present for a small number of respondents, but we do not find systematic evidence of a greater bias toward under- or overreporting for parent-report versus self-report data (Finkelhor, Hamby, Ormrod, & Turner, 2005). Previous analyses have also demonstrated good respondent comprehension, construct validity, and test–retest reliability for JVQ items (Finkelhor et al., 2005). Researchers who have looked at this issue have found that while there is likely to be some underreporting in parent and self-reports, these reports are likely to be more valid than official reports that greatly underestimate victimization experiences (Hardt & Rutter, 2004; Kendall-Tackett & Becker-Blease, 2004; Straus, 1992).

The analyses presented in this article are based on 20 items covering four domains of victimization—sexual assault, maltreatment, witnessing family violence, other major violence—the screener questions for each module are available in Appendix A of Turner, Finkelhor, and Ormrod (2006). The complete text of the questions is available in Hamby et al. (2004). Participants were asked about each event in the past year. If no event was reported for the past year, participants were asked if the event had ever happened in their lifetime. From these data, we calculated the number of types of events reported across participants' lifetimes.

Nonvictimization adversity. The nonvictimization adversity module of the JVQ includes 15 stressful events and chronic stressors, including an item assessing exposure to a disaster. For the purposes of these analyses, we include only the 14 nondisaster items in the adversity score.

Lifetime and past-year disaster exposure. Disaster exposure was measured using the following item, "In your (your child's) whole life, were you (was she or he) ever in a VERY BAD fire, explosion, flood, tornado, hurricane, earthquake or other disaster?" For those respondents who answered yes, follow-up questions assessed how many times, whether the most recent event happened in the past year, what type of disaster was experienced, the child's age at the time of the disaster, and if the child knew anyone who was injured in the disaster.

Current mental health symptoms. Mental health symptoms were measured using the anxiety, depression, and aggression subscales of the Trauma Symptom Checklist for Young Children (Briere et al., 2001; for children aged 2–9) and the Trauma Symptom Checklist for Children (Briere, 1996; for children aged 10–17). Both instruments assess reactions to a wide range of nonspecific stressful events across several domains. Participants report on how frequently they (or, for parents of 2- to 9-year-olds, their child) had experienced a list of behaviors, thoughts and feelings on a 4-point scale (0 = *not at all*, 4 = *very often*) over the past month. In this sample, approximately 5% of respondents scored in the subclinical or clinical range for anxiety, depression, or aggression. Because of the small number of extreme scores, we used continuous measures of mental health symptoms in our analyses.

Analysis Plan

We used 2002 Census estimates (U.S. Census Bureau, 2000) to apply poststratification weights to adjust for the underrepresentation of Blacks and Hispanics in our sample relative to the national population. Weights also adjusted for the probability of selection due to the number of children in the household and the fact that only one child per household was selected to participate. Unless otherwise noted, analyses were conducted with weighted data. Due to the potential for method variance between the younger (parent report) and older (self-report) groups, data for the younger and older groups were analyzed separately. Within each group, age was positively associated with disaster and victimization exposure. Thus, where indicated, age is included as a covariate.

In tests of group differences, children who had experienced any incident within a victimization domain (sexual assault, maltreatment, witness domestic violence, other major violence) were included in that category of victimization and in the category of "any victimization."

For Hypothesis 1, we ran frequency counts for children reporting disaster and victimization experiences.

To test for associations between experiencing a disaster and victimization (Hypothesis 2), we conducted chi-square analyses to examine associations between lifetime disaster, victimization, adversity experience. We also used partial correlation to test for associations among disaster exposure, adversity, any victimization, sexual victimization, maltreatment, domestic violence, and other major violence, controlling for age.

For Hypothesis 3, we ran regression analyses with victimization and control variables predicting mental health symptoms among participants who reported experiencing a disaster.

For Hypothesis 4, we used analysis of covariance (ANCOVA) to compare mental health symptoms among four groups: no disaster or victimization, disaster only, victimization only, and both disaster and victimization, controlling for demographic factors. Bonferroni-corrected contrasts were used to identify significant differences among the four groups.

Regression analyses were used to identify predictors of mental health symptoms when group differences indicated greater symptoms among disaster victims relative to peers who did not experience disaster. In these analyses, victimization was calculating by summing the number of reported incidents across the four domains assessed. The theoretical range was 0–20 and participants reported between 0 and 12 incidents ($M = 0.82$, $SD = 1.52$). Adversity scores were obtained by giving a score of 1 for each type of adversity that the child had ever experienced and summing to form a continuous score. Race and ethnicity was coded into four groups: White and non-Hispanic, Black and non-Hispanic, Other Race and non-Hispanic, and Hispanic. Family structure was coded into three groups: single parent, stepparent household, and two biological or adoptive parents. To construct an index of SES, income and parental education scores were separately standardized, summed and then restandardized. When either income or education data were missing, the SES score is based on the variable available.

Results

Table 1 shows demographic variables for the sample.

How Many Children Are Affected by Disasters in the United States?

Based on analyses with weighted data, 281 or 13.9% of children and teens had been in a disaster during their lifetime, and 4.1% of the sample experienced a disaster in the past year. By comparison, 13.7% reported sexual victimization, 20.0% reported maltreatment, 10.6% reported witnessing family violence, and 17.3% reported witnessing other major violence within their lifetimes. In total, 36.4% reported experiencing at least one victimization incident. Age at the time of disaster ranged from < 1 year old through 16 years old.

Of those who had experienced a disaster in their lifetime, 24.6% reported experiencing more than one. With respect to the most recent event, 12.8% had been in a fire, 26.7% in a tornado, 14.6% in a fire, 24.2% in a hurricane, and 17.8% in an earthquake. Eight participants indicated other disasters including bad ice and wind storms, the Oklahoma

Table 1
Demographics by Age (Unweighted Data)

	2- to 9-year-olds (*M* age = 5.4 years, *SD* = 2.4)		10- to 17-year-olds (*M* age = 13.8 years, *SD* = 2.2)	
	n (1,030)	%	*n* (1,000)	%
Female sex	518	50.3	491	49.1
Child race/ethnicity[a]				
White, non-Hispanic	760	73.8	776	77.6
Black, non-Hispanic	117	11.4	103	10.3
Other, non-Hispanic	40	3.9	31	3.1
Hispanic, any race	100	9.7	81	8.1
U.S. region[b]				
Northeast	224	21.7	217	21.7
Midwest	289	28.1	247	24.7
South	327	31.7	335	33.5
West	190	18.4	201	20.1
Metropolitan area[c]				
Rural	191	18.5	279	27.9
Town	230	22.3	219	21.9
Small city	216	21.0	162	16.2
Large city suburb	216	21.0	199	19.9
Large city	168	16.3	132	13.2
Parents in home				
Two parents (bio or adoptive)	778	75.5	641	64.1
One parent + partner	59	5.7	126	12.6
Single/other	193	18.7	233	23.3
Highest parent education[d]				
HS/Vo-tech, or less	250	24.3	273	27.3
Some college/college graduate	576	55.9	516	51.6
Graduate school	203	19.7	211	21.1
Household income (per year)[e]				
Under $20,000	98	9.5	100	10.0
$20,000–$50,000	376	36.5	317	31.7
More than $50,000	464	45.0	469	46.9

Note. Except for region of the United States, all demographic data were provided by the caregiver.
[a] Respondents indicated "unknown" race/ethnicity information for 13 children aged 2–9 and 9 children aged 10–17. [b] U.S. regions correspond to Census data and were derived from telephone area codes. [c] Metropolitan areas were defined as follows: rural = population < 20,000; town = 20,000–100,000; large city = population of 300,000 or more. Data were not available for 9 children aged 2–9 and 9 children aged 10–17. [d] Highest parent education is the highest level of education obtained by any parent in the home and was reported by the caregiver. Data were not available for the caregivers of 1 child aged 2–9. [e] Caregivers declined to provide household income information for 92 children aged 2–9 and 114 children aged 10–17.

City bombing, and World Trade Center attacks. Nine participants (.4%) reported being injured in the disaster.

Is Risk for Disaster Exposure Associated With Risk for Exposure to Other Stressful Events?

Table 2 presents the percentage of participants who experienced victimization and other adversity, by disaster exposure history and age group. These results do not control for age within each age group.

Partial correlation was used to examine relations between lifetime disaster exposure and history of victimization and adversity, controlling for age. Contrary to Hypothesis 2, for 2- to 9-year-olds, no correlation was found between lifetime disaster experience and lifetime sexual victimization ($r = .03$, $p = .31$), witnessing domestic violence ($r = .02$, $p = .45$), other major violence ($r = -.03$, $p = .30$), or any kind of victimization ($r = .02$, $p = .53$) when age was controlled. However, experiencing a lifetime disaster was positively associated with experiencing maltreatment ($r = .06$, $p = .04$) and adversity ($r = .08$, $p = .1$) with age controlled.

Supporting Hypothesis 2, among 10- to 17-year-olds, correlation analyses controlling for age showed a positive relation between lifetime disaster exposure and lifetime sexual victimization ($r = .86$, $p < .001$), maltreatment ($r = .29$, $p < .001$), and adversity ($r = .13$, $p < .001$). There was no correlation with witnessing domestic violence ($r = -.002$, $p = .48$). Contrary to predictions, a negative relation emerged between lifetime disaster exposure and major violence ($r = -.15$, $p < .001$) and any kind of victimization ($r = -.16$, $p < .001$) with age controlled.

Is Victimization Associated With Mental Health Among Lifetime Disaster Victims?

Controlling for sex, race and ethnicity, parent makeup, age, and SES, victimization but not adversity was associated with mental health symptoms (see Table 3). Specifically, among 2- to 9-year-olds, victimization predicted depression scores. Among 10- to 17-year-olds, victimization predicted depression and anger/aggression scores.

How Do Mental Health Symptoms Compare for Children Who Have Experienced Disasters, Victimization, Both Disaster and Victimization Events, and Neither Disaster Nor Victimization Events?

As shown in Table 4, among 2- to 9-year-olds, lifetime disaster and victimization exposure was significantly associated with anxiety, depression, and aggression in a model that included sex, age, race and ethnicity, parent makeup, and SES. Marginal means and Bonferroni-corrected contrasts comparing participants who experienced neither disaster nor victimization, disaster only, victimization only, or both disaster and victimization are presented in Table 5 and Figure 1. Those who experienced victimization only had higher scores than those who experienced neither disaster nor victimization or disaster only. Two- to 9-year-olds who experienced disaster only had unexpectedly low scores that did not differ significantly from those who had experienced neither a disaster nor victimization.

Table 2

Estimated Percent of Disaster Victims Who Also Experienced Victimization and Other Adversity

	Sex victim	Maltreatment	Witness family violence	Major violence	Any victim	Adversity
2- to 9-year-olds (*n*)						
Disaster victims (75)	2.7	8.0	6.7	13.3	24	18.7_a
No disaster reported (933)	4.2	15.1	9.4	8.9	25.6	64.3_a
10- to 17-year-olds (*n*)						
Disaster victims (207)	30.4_b	35.7_c	14.0	39.1_d	64.7_e	99.0
No disaster reported (809)	21.2_b	22.6_c	11.4	21.8_d	42.6_e	88.9

Note. Chi-square tests revealed statistically significant association between those cells marked with identical subscripts.
${}_a\chi^2(1, N = 1{,}008) = 8.91$, $p = .002$. ${}_b\chi^2(1, N = 1{,}014) = 7.80$, $p = .01$. ${}_c\chi^2(1, N = 1{,}014) = 15.28$, $p < .001$. ${}_d\chi^2(1, N = 1{,}014) = 26.68$, $p < .001$. ${}_e\chi^2(1, N = 1{,}014) = 32.90$, $p < .001$.
Because data are weighted, observed cell counts are not whole integers. Rounding within cells resulted in $n = 282$ disaster victims for the sexual victimization and witnessing family violence. For maltreatment, major violence, and any victimization analyses, disaster victims $n = 281$. Four participants indicated they were "not sure" if the child/teen had been in a disaster and are not included in this analysis.

Table 3

Multivariate and Univariate Analysis of Variance for Mental Health Symptoms Among Disaster Victims

	2- to 9-year-olds						10- to 17-year-olds					
	Multivariate tests		Univariate tests				Multivariate tests		Univariate tests			
	df	*F*	*df*	Anx.	Dep.	Ang.	*df*	*F*	*df*	Anx.	Dep.	Ang.
Sex (female = 1)	(3, 54)	0.81	(1, 56)	0.20	0.88	0.48	(3, 168)	0.418	(1, 180)	0.76	0.07	0.02
Race/ethnicity	(9, 168)	1.52	(3, 56)	1.34	2.63	2.94*	(9, 510)	1.64	(3, 180)	2.83*	4.50**	3.31*
Parent makeup	(6, 110)	1.41	(2, 56)	0.16	1.28	2.07	(6, 338)	1.95	(2, 180)	3.87*	1.09	3.63*
SES	(3, 54)	3.01*	(1, 56)	8.34**	0.06	2.04	(3, 168)	0.057	(1, 180)	0.01	0.05	0.02
Age	(3, 54)	1.03	(1, 56)	0.10	0.53	2.78	(3, 168)	0.73	(1, 180)	1.17	0.86	2.21
Adversity	(3, 54)	1.58	(1, 56)	1.89	3.67	0.06	(3, 168)	1.44	(1, 180)	0.35	2.03	2.86
Victimization	(3, 54)	13.21***	(1, 56)	2.72	38.44***	1.31	(3, 168)	3.58*	(1, 180)	3.40	10.47**	4.37*

Note. Multivariate *F* ratios were generated from Pillai's statistic. SES = socioeconomic status.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 4

Multivariate and Univariate Analysis of Variance for Mental Health Symptoms: Comparisons by Disaster and Victimization History

	2- to 9-year-olds						10- to 17-year-olds					
	Multivariate tests		Univariate tests				Multivariate tests		Univariate tests			
	df	*F*	*df*	Anx.	Dep.	Ang.	*df*	*F*	*df*	Anx.	Dep.	Ang.
Sex (female = 1)	(3, 884)	0.81	(1, 886)	4.23*	0.34	4.79*	(3, 884)	3.73*	(1, 886)	.87	2.33	1.12
Race/ethnicity	(9, 2658)	1.52	(1, 886)	2.77*	6.26***	3.59*	(9, 2658)	2.32*	(1, 886)	2.91*	1.37	2.11
Parent makeup	(6, 1770)	1.41	(1, 886)	1.00	7.40**	3.95*	(6, 1770)	1.53	(1, 886)	0.01	1.68	1.50
SES	(3, 884)	3.07*	(1, 886)	0.10	0.30	0.04	(3, 884)	6.54***	(1, 886)	3.40	13.64***	15.30***
Age	(3, 884)	1.03	(1, 886)	0.36	9.30**	78.29***	(3, 884)	5.65**	(1, 886)	2.23	0.59	3.69
Victimization/disaster Group[a]	(9, 2658)	13.21***	(1, 886)	25.87***	32.39***	9.92***	(9, 2658)	14.64***	(1, 886)	37.62***	34.14***	37.53***

Note. Multivariate *F* ratios were generated from Pillai's statistic. SES = socioeconomic status.
[a] Victimization groups are: (a) no reported victimization or disaster, (b) victimization history without disaster exposure, (c) disaster exposure without victimization history, and (d) both victimization and disaster exposure.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Also shown in Table 5, as well as in Figure 2, are mental health scores by disaster and victimization group among 10- to 17-year-olds. In interpreting Figure 2, it is important to consider the variance in standard errors of the estimated marginal means. For each mental health subscale, disaster survivors' standard errors of the estimated marginal mean symptom score are greater than for the other three groups. This contributes to the fact that contrasts between scores for disaster survivors and the other three groups in many cases failed to reach traditional levels of significance. The mean mental health symptoms scores for lifetime disaster only survivors were in the predicted direction, that is, higher than those who had experienced neither a victimization nor a disaster, and lower than those who had experienced either victimization only or victimization and disaster. Consistent with Hypothesis 4, 10- to 17-year-olds who experienced neither a disaster nor victimization had the lowest anxiety, depression, and aggression scores.

Discussion

To the best of our knowledge, this article provides the first prevalence and incidence estimate of disasters in a nationally representative sample of U.S.

Table 5

Estimated Marginal Means for Mental Health Scores by Age and Disaster and Victimization History

	Anxiety		Depression		Anger/aggression	
	Estimated marginal mean	Standard error	Estimated marginal mean	Standard error	Estimated marginal mean	Standard error
2- to 9-year-olds (*n*)						
Neither (614)	2.73_{a}	0.21	1.81_{cd}	0.13	2.80_{h}	0.22
Victimization only (217)	4.00_{ab}	0.25	2.97_{cef}	0.16	5.44_{hi}	0.27
Disaster only (54)	2.59_{b}	0.44	1.43_{eg}	0.28	2.63_{i}	0.48
Both (217)	4.64_{b}	0.85	4.42_{dfg}	0.54	3.91	0.92
10- to 17-year-olds (*n*)						
Neither (409)	1.92_{jkl}	0.24	1.70_{mno}	0.25	2.22_{pqr}	0.29
Victimization only (308)	4.66_{j}	0.24	4.40_{m}	0.25	5.44_{p}	0.29
Disaster only (59)	2.88_{k}	0.48	3.13_{n}	0.49	4.63_{q}	0.57
Both (122)	4.50_{l}	0.34	4.04_{o}	0.35	5.19_{r}	0.41

Note. Bonferroni-corrected contrasts indicated that means with identical subscripts were significantly different ($p < .05$) from each other.

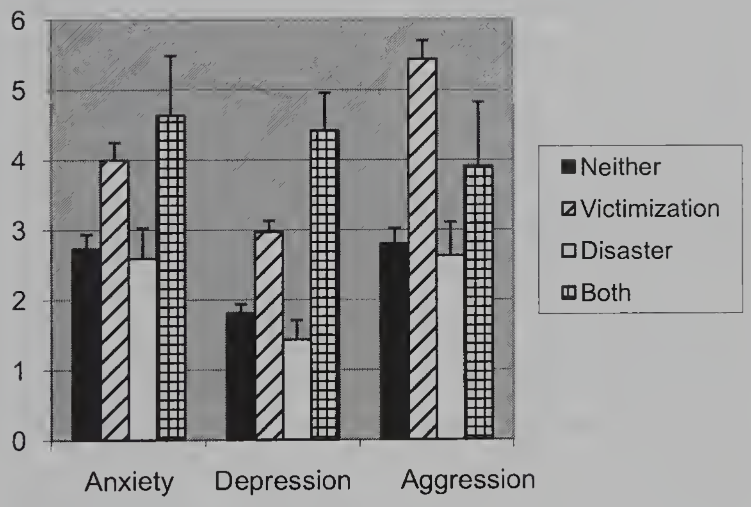


Figure 1. Marginal means for 2- to 9-year-olds' mental health symptoms by victimization and disaster exposure.

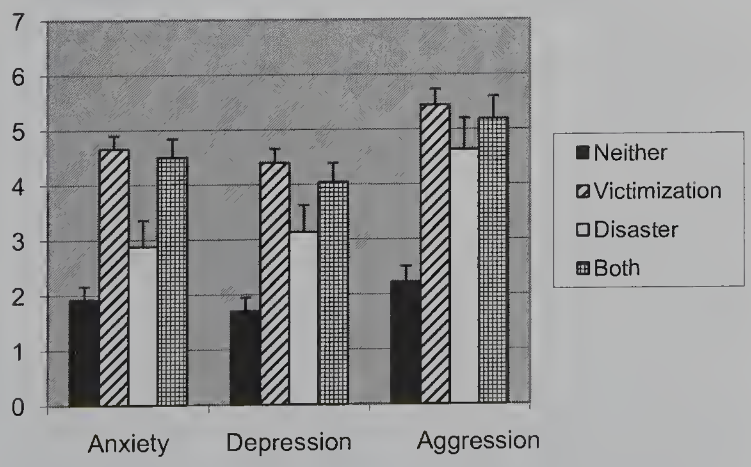


Figure 2. Marginal means for 10- to 17-year-olds' mental health symptoms by victimization and disaster exposure.

children. While the number of studies documenting the effects of disasters on children is increasing, it is important to know how many children experience disasters in order to understand the magnitude of the effects of disasters nationwide.

It was surprising that only 2 of 70 past-year disaster victims reported receiving counseling. We had expected more children to have received counseling because disasters tend to be relatively publicly known, nonstigmatized, and elicit community response. Some children may have received counseling following a disaster, but it was not reported because of the wording of our question that asked about counseling ''for emotional or behavioral problems.'' It is quite possible that some participants did not consider disaster-related counseling to be ''for emotional or behavioral problems.''

We found an association between exposure to disasters, maltreatment, and adversity (controlling for age) among both age groups. Risk for disaster exposure shares some risk factors with risk for maltreatment and adversity. It is also possible that disaster exposure increases children's risk for subsequent maltreatment and adversity. Some of the adversities measured could have been directly related to a disaster (e.g., being homeless, knowing someone who died, seeing a dead body). Others are less likely to be related (e.g., being teased about one's appearance).

Regardless of the overlap in exposure to disasters and victimization, mental health professionals would benefit from knowing if previous or concurrent victimization is a significant predictor of mental health among those who have experienced a disaster. We found that this is the case. Among lifetime disasters survivors, victimization history predicted 2- to 9-year-olds' depression scores, and 10- to 17-year-olds' depression and anxiety scores. We lacked the information to know if victimization is associated with children acutely exposed to disaster, but the association between lifetime exposure to disaster and victimization suggests the need for screening for victimization history among young disaster victims when treating mental health symptoms.

Finally, we compared mental health symptoms among children who have experienced disaster only, victimization only, neither disaster nor victimization, or both disaster and victimization. A surprising finding affected the interpretation of the hypothesized effect: Two- to 9-year-olds who experienced both disasters and victimization and 10- to 17-year-olds who experienced disasters only had more variable mental health scores than peers in the other three groups. This effect should be tested in future research. If this result stands, it suggests a need for careful screening to differentiate those children who are at risk for negative mental health outcomes.

The possibility of method bias is introduced by having parents report on behalf of young children while older youth provided self-reports. The decision to use different reporters for younger versus older youth was made to reduce bias that is known to exist when parents are asked to report on their older children's experiences and mental health. Based on the available evidence, this procedure offered the most accurate data available given the subject matter. Nevertheless, we do not have retrospective reports from participants later in life, nor do we have corroborative reports from other sources. Additional clinical measures (e.g., structured clinical interviews) would have yielded additional information about the diagnosis of mental disorders in this sample.

Despite many strengths of this study, there are some limitations to discuss. Because we sampled households with telephones, the sample probably underrepresented low-income and homeless families. We asked about victimization, disaster, and adversity prior to asking about mental health symptoms. It is possible that the sequence of questions led to higher reports of mental health symptoms. On the other hand, youth experiencing depression and anxiety symptoms may have been more likely to recall and report bad events, inflating the association between negative past events and mental health. The retrospective nature of reports could lead to underreporting, especially of sensitive questions about abuse. Because many abuse events are not reported to authorities, we believe self-reports are an important method for collecting these data, but underreporting could have led to underestimating associations between some victimization experiences and mental health. We do not, in these data, have the ability to know the sequence of disaster exposure, victimization, other adversity and onset of mental health symptoms. Longitudinal, prospective data would better address the important question of whether being in a disaster leads to other adversities or even puts children at risk for victimization.

In addition to addressing these limitations, future research may also assess several factors that may well mediate or moderate the association between disaster exposure, victimization, and mental health symptoms, such as level of fear, loss of property, homelessness, lack of food and water, and separation from caregivers. In fact, some of these factors may be more related to certain types of posttraumatic stress disorder or other mental health symptoms than others (Briere & Elliott, 2000). Outcomes other than mental health symptoms should be considered in a representative sample as well, including changes (positive and negative) in relationships, school performance, and parenting. Overall, the results point to important interactions among various kinds of stressful events and a need to continue to study various forms of stressful events simultaneously.

References

Bajos, N., Spira, A., Ducot, B., & Messiah, A. (1992). Analysis of sexual behavior in France (ACSF): A comparison between two modes of investigation: Telephone survey and face-to-face survey. *AIDS, 6*, 315–323.

Bermack, E. (1989). Effects of telephone & face-to-face communication on rated extent of self-disclosure by female college students. *Psychological Reports, 65*, 259–267.

Brick, J., Waksberg, J., Kulp, D., & Starer, A. (1995). Bias in list-assisted telephone samples. *Public Opinion Quarterly, 59*, 218–235.

Briere, J. (1996). *Trauma Symptom Checklist for Children: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Briere, J., & Elliott, D. (2000). Prevalence, characteristics and long-term sequelae of natural disaster exposure in the general population. *Journal of Traumatic Stress, 13*, 661–679.

Briere, J., Johnson, K., Bissada, A., Damon, L., Crouch, J., Gil, E., et al. (2001). The Trauma Symptom Checklist for Young Children (TSCYC): Reliability and association with abuse exposure in a multi-site study. *Child Abuse & Neglect, 25*, 1001–1014.

Bromet, E. J., Goldgaber, D., Carlson, G., Panina, N., Golovakha, E., Gluzman, S. F., et al. (2000). Children's well-being 11 years after the Chornobyl catastrophe. *Archives of General Psychiatry, 57*, 563–571.

Bureau of Justice Statistics. (various years). *National crime victimization survey*. Washington, DC: Bureau of Justice Statistics.

Centers for Disease Control (2005). Effects of Hurricane Katrina on children's blood lead levels. Retrieved June 10, 2008, from http://emergency.cdc.gov/disasters/hurricanes/katrina/leadkatrina.asp

Cornely, P., & Bromet, E. J. (1986). Prevalence of behavior problems in three-year-old children living near Three Mile Island: A comparative analysis. *Journal of Child Psychology and Psychiatry, 27*, 489–498.

Curtis, T., Miller, B. C., & Berry, E. H. (2000). Changes in reports and incidence of child abuse following natural disasters. *Child Abuse & Neglect, 24*, 1151–1162.

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 2301–2306.

Czaja, R. (1987). Asking sensitive behavioral questions in telephone interviews. *International Quarterly of Community Health Education, 8*, 23–32.

Dirkzwager, A. J. E., Kerssens, J. J., & Yzermans, C. J. (2006). Health problems in children and adolescents before and after a man-made disaster. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*, 94–103.

Dong, M., Anda, R. F., Felitti, V. J., Dube, S. R., Williamson, D. F., Thompson, T. J., et al. (2004). The interrelatedness of multiple forms of childhood abuse, neglect, and household dysfunction. *Child Abuse & Neglect, 28*, 771–784.

Durkin, M. S., Khan, N., Davidson, L. L., Zaman, S. S., & Stein, Z. (1993). The effects of a natural disaster on child behavior: Evidence for posttraumatic stress. *American Journal of Public Health, 83*, 1549–1553.

Finkelhor, D., Hamby, S. L., Ormrod, R., & Turner, H. (2005). The Juvenile Victimization Questionnaire: Reliability, validity, and national norms. *Child Abuse & Neglect, 29*, 383–412.

Finkelhor, D., & Kendall-Tackett, K. (1997). A developmental perspective in the childhood impact of crime, abuse and violent victimization. In D. Cicchetti & S. L. Toth (Eds.), *Developmental perspectives on trauma: Theory, research and intervention* (Vol. 8, pp. 1–32). Rochester, NY: University of Rochester Press.

Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). Polyvictimization and trauma in a national longitudinal cohort. *Development and Psychopathology, 19*, 149–166.

Goldberg, L. R., & Freyd, J. J. (2006). Self-reports of potentially traumatic experiences in an adult community sample: Gender differences and test-retest stabilities of the items in a brief betrayal-trauma survey. *Journal of Trauma & Dissociation, 7*, 39–63.

Green, B. L., Grace, M. C., Vary, M. G., Kramer, T. L., Gleser, G. C., & Leonard, A. C. (1994). Children of disaster in the second decade: A 17-year follow-up of buffalo creek survivors. *Journal of the American Academy of Child & Adolescent Psychiatry, 33*, 71–79.

Green, B. L., Korol, M., Grace, M. C., Vary, M. G., Leonard, A. C., Gleser, G. C., et al. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*, 945–951.

Green, B. L., & Solomon, S. D. (1995). The mental health impact of natural and technological disasters. In J. R. Freedy & S. E. Hobfoll (Eds.), *Traumatic stress: From theory to practice*. (pp. 163–180). New York: Plenum

Hamby, S. L., Finkelhor, D., Ormrod, R., & Turner, H. (2004). *The Juvenile Victimization Questionnaire (JVQ): Administration and scoring manual*. Durham, NH: Crimes Against Children Research Center.

Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry, 45*, 260–273.

Hausman, A. J., Spivak, H., Prothrow-Stith, D., & Roeber, J. (1992). Pattern of teen exposure to a community-based violence prevention project. *Journal of Adolescent Health, 13*, 668–675.

Healthfinder.gov. (2008). Boys Town National Hotline. Retrieved January 10, 2009, from http://www.healthfinder.gov/orgs/HR3689.htm

Keenan, H. T., Marshall, S. W., Nocera, M. A., & Runyan, D. K. (2004). Increased incidence of inflicted traumatic brain injury in children after a natural disaster. *American Journal of Preventive Medicine, 26*, 189–193.

Kendall-Tackett, K., & Becker-Blease, K. (2004). The importance of retrospective findings in child maltreatment research. *Child Abuse & Neglect, 28*, 723–727.

Landrigan, P. J., Forman, J., Galvez, M., Newman, B., Engel, S. M., & Chemtob, C. (2008). Impact of September 11 World Trade Center disaster on children and pregnant women. *Mount Sinai Journal of Medicine, 75*, 129–134.

Lund, L., & Wright, W. E. (1994, March). *Mitofsky-Waksberg vs. screened random digit dial: Report on a comparison of the sample characteristics of two RDD survey designs*. Paper presented at the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System conference, Atlanta, GA.

Marin, G., & Marin, B. (1989). A comparison of three interviewing approaches for studying sensitive topics with Hispanics. *Hispanic Journal of Behavioral Sciences, 11*, 330–340.

Masozeraa, M., Bailey, M., & Kerchner, C. (2007). Distribution of impacts of natural disasters across income groups: A case study of New Orleans. *Ecological Economics, 63*, 299–306.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. (2002). 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry: Interpersonal and Biological Processes*, *65*, 207–239.

Rath, B., Donato, J., Duggan, A., Perrin, K., Bronfin, D. R., Ratard, R., et al. (2007). Adverse health outcomes after Hurricane Katrina among children and adolescents with chronic conditions. *Journal of Health Care for the Poor and Underserved*, *18*, 405–417.

Saylor, C. F., Swenson, C. C., & Powell, P. (1992). Hurricane Hugo blows down the broccoli: Preschoolers' post-disaster play and adjustment. *Child Psychiatry & Human Development*, *22*, 139–149.

Stein, B. D., Jaycox, L. H., Elliott, M. N., Collins, R., Berry, S., Marshall, G. N., et al. (2004). The emotional and behavioral impact of terrorism on children: Results from a national survey. *Applied Developmental Science*, *8*, 184–194.

Stein, B. D., Jaycox, L. H., Kataoka, S., Rhodes, H. J., & Vestal, K. D. (2003). Prevalence of child and adolescent exposure to community violence. *Clinical Child & Family Psychology Review*, *6*, 247–264.

Straus, M. A. (1992). The conflict tactic scales and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8145 families* (pp. 49–73). New Brunswick, NJ: Transaction.

Turner, H. A., Finkelhor, D., & Ormrod, R. (2006). The effect of lifetime victimization on the mental health of children and adolescents. *Social Science & Medicine*, *62*, 13–27.

U.S. Census Bureau. (2000). *(NP-D1-A) projections of the resident population by age, sex, race, and Hispanic origin: 1999–2100*. Washington, DC: Population Projections Program—Population Division.

Waksberg, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, *361*, 40–46.

Weeks, M. F., Kulka, R. A., Lessler, J. T., & Whitmore, R. W. (1983). Personal versus telephone surveys from collecting household health data at the local level. *American Journal of Public Health*, *73*, 1389–1394.

Weems, C. F., Pina, A. A., Costa, N. M., Watts, S. E., Taylor, L. K., & Cannon, M. F. (2007). Predisaster trait anxiety and negative affect predict posttraumatic stress in youths after Hurricane Katrina. *Journal of Consulting and Clinical Psychology*, *75*, 154–159.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1053–1076

Unpacking Trauma Exposure Risk Factors and Differential Pathways of Influence: Predicting Postwar Mental Distress in Bosnian Adolescents

Christopher M. Layne
UCLA/Duke University National Center for Child Traumatic Stress, University of California, Los Angeles

Joseph A. Olsen
Brigham Young University

Aaron Baker
University of California, Los Angeles

John-Paul Legerski
University of North Dakota

Brian Isakson
Center for Rural and Community Behavioral Health, University of New Mexico

Alma Pašalić
Sarajevo University Medical Center

Elvira Duraković-Belko and
Nermin Đapo
University of Sarajevo

Nihada Ćampara
Travnik Hospital, Bosnia

Berina Arslanagić
UNICEF Mongolia

William R. Saltzman and
Robert S. Pynoos
UCLA/Duke University National Center for Child Traumatic Stress, University of California, Los Angeles

Methods are needed for quantifying the potency and differential effects of risk factors to identify at-risk groups for theory building and intervention. Traditional methods for constructing war exposure measures are poorly suited to "unpack" differential relations between specific types of exposure and specific outcomes. This study of 881 Bosnian adolescents compared both *common factor–effect indicator* (using exploratory factor analysis) versus *composite causal–indicator* methods for "unpacking" dimensions of war exposure and their respective paths to postwar adjustment outcomes. The composite method better supported theory building and most intervention applications, showing how multitiered interventions can enhance treatment effectiveness and efficiency in war settings. Used together, the methods may unpack the elements and differential effects of "caravans" of risk and promotive factors that co-occur across development.

In the aftermath of wars and other catastrophic events, questions concerning whom and how to help—that is, *how to best identify at-risk groups and allocate scarce mental health resources*—confront every practitioner and agency that strives to promote the recovery of affected youths and families. Efforts to address these questions have contributed to our growing knowledge base by improving the method-

---

Financial support was provided by UNICEF Bosnia & Herzegovina, the BYU Family Studies Center, the David M. Kennedy Center for International Studies, the Bing Fund, and Tony Bennett. The authors thank Preston Finley, James Wu, Ann Anderson, and Jenifer Alonso for assisting with manuscript preparation. Special thanks to Theresa Betancourt, Alan Steinberg, Sarah Ostrowski, Harolyn Belcher, and two anonymous reviewers for their generous and insightful comments on previous drafts of this manuscript.

Correspondence concerning this article should be addressed to Christopher M. Layne, UCLA National Center for Child Traumatic Stress, 11150 W. Olympic Blvd., Suite 650, Los Angeles, CA 90064. Electronic mail may be sent to cmlayne@mednet.ucla.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0004

ological rigor of disaster research (Norris, Galea, Friedman, & Watson, 2006), building theory by identifying the correlates and sequelae of war exposure (e.g., Dyregrov, Gupta, Gjestad, & Mukanoheli, 2000; Macksoud & Aber, 1996), and field-testing interventions (e.g., Layne et al., 2008). These studies have identified a broad range of types of distress in war-exposed youths, including posttraumatic stress disorder (PTSD), depression and other internalizing disorders, externalizing behaviors, impaired school performance, and disturbances in personality formation and moral development (see Barenbaum, Ruchkin, & Schwab-Stone, 2004; Entholt & Yule, 2006; Fayyad, Karam, Karam, Tabet, & Ghosn, 2004; Shaw, 2003).

Although the literature identifies a general dose–response relation between the magnitude of war exposure and youth postwar distress, studies are providing growing evidence of *differential relations* between specific trauma-related risk factors and specific posttraumatic adjustment outcomes (Layne, Beck, et al., 2009). In their review, Fayyad et al. (2004) note that for at least half of war-exposed youths, PTSD reactions decrease on their own in the postwar aftermath, whereas other youths continue to meet full PTSD diagnostic criteria or experience delayed-onset stress reactions. The persistence of these reactions is linked to both the degree of severity of initial trauma exposure as well as the presence of postwar family stressors, such as maternal dysfunction and poverty. Such findings underscore the influential roles played not only by the specific *types* of trauma and loss to which youths are exposed but also of associated moderators and mediators of youth postwar adjustment (see Miller & Rasmussen, 2010). For example, evidence that having one's parent killed and the disappearances of loved ones are especially traumatogenic (Shaw, 2003) suggests that war-related causal risk factors may differ significantly in their potency. In addition, a gender-moderated effect is evident in findings that war-exposed girls generally manifest higher levels of anxiety and mood symptoms than boys, whereas boys manifest higher rates of externalizing behavior (Pat-Horenczyk et al., 2009; Shaw, 2003). The importance of mediators is underscored by evidence that different war-related causal risk factors may exert their effects on postwar outcomes through different pathways of influence. For example, Entholt and Yule (2006) report that PTSD symptoms in refugee youths are more strongly linked to earlier war trauma and subsequent resettlement strain, whereas depression symptoms are more strongly linked to recent stressful life events and current maternal mental health. Taken together, these studies constitute important early progress in understanding the mechanisms and processes that underlie the etiology and clinical course of youth psychosocial maladjustment following war-related trauma and loss.

However, the progress made in the war and disaster literature in identifying the pathways through which specific types of trauma exposure differentially influence posttraumatic psychosocial adjustment is impeded by four methodological practices. As discussed by Netland (2001, 2005) and Barenbaum et al. (2004), these practices include: (a) the inappropriate use of the *common latent factor with effect indicators model* (i.e., factor analysis) to derive the dimensionality of trauma exposure measures; (b) inadequate event sampling in scale construction (e.g., selecting a best *sample* of ''internally consistent'' events vs. a comprehensive event *census*), including the practice of importing existing trauma exposure measures into new settings that differ markedly from those for which the measures were created; (c) the use of total-scale summative scoring (e.g., creating a unit-weighted ''total war exposure'' score that intermixes many different types of trauma); and (d) the use of scale reliability indices (e.g., Cronbach's alpha; Cronbach, 1951) that are incompatible with the psychometric assumptions or particular applications of the measurement model being used.

Consequently, although the literature provides compelling evidence that war exposure is generally harmful to youth adjustment, its ability to explain *which* types of exposure are most harmful, for *which* outcomes, for *whom*, *how* they transmit their effects, and *how* this knowledge should guide intervention, is much more limited (Layne, Warren, Watson, & Shalev, 2007). Accordingly, although noting a general dose–response relation between trauma exposure and mental health, Barenbaum et al. (2004) offer the caveat that different types of trauma relate differentially to PTSD and other outcomes, such that some traumatic events have only moderate or negligible long-term effects whereas others are devastating. The authors note that the literature is largely unable to explain how or why this occurs. Such shortcomings recently led an APA task force to conclude that although some posttrauma *indicators of risk* have been identified, no method exists for accurately predicting whether specific youths will recover on their own or instead require some intervention. The task force called for the development of ''well-validated risk assessment tools that can be feasibly implemented *in diverse settings and for diverse traumatic events and that will help identify the high-risk youth and families who are in need of* *clinical services*" (APA Presidential Task Force, 2008, p. 6, italics added).

Developing risk assessment methods and tools for quantifying the potency and differential effects of specific types of trauma exposure that are valid and useful across diverse settings carries great promise for advancing the field, especially through building theory, enhancing risk screening and intervention, and guiding social policy (Kraemer et al., 1999). Specifically, a method that aggregates into the same test dimension *equifinal* causal risk factors (i.e., factors that lead to the same causal consequences; Richters, 1997) will help to quantify the differential effects of specific types of risk factors on specific outcomes. This practice can build theory by clarifying the etiological significance of specific types of trauma and loss for specific outcomes and the differential pathways of influence that interconnect them. In turn, clarifying *who is at risk for what outcome* will improve risk screening and triage by assigning groups with different types of trauma and loss (e.g., traumatic bereavement) to intervention modules that target those types of distress for which members are at greatest risk (e.g., maladaptive grief reactions). Equally valuable, knowledge of key pathways of influence can improve the effectiveness and efficiency of interventions by identifying "high-value" targets for intervention. This will guide treatments in addressing not only "endpoint" distress—related *outcomes* (e.g., PTSD, depression, traumatic grief), but also *key causal contributors to* those outcomes (e.g., trauma and loss reminders).

Given these advantages, we propose that it is a propitious time to reevaluate the ways in which war exposure is measured. Much can be gained by supplementing the traditional test construction method of factor analysis and related approaches (e.g., Dyregrov et al., 2000) that "unpack" trauma exposure into dimensions that *model the structure of event co-occurrences*, with a method that "unpacks" exposure into dimensions that *model the occurrences of specific types of trauma and loss and their differential links to psychosocial consequences*. As part of this reevaluation, candidate methods for constructing measures of trauma exposure should be evaluated according to their promise for strengthening research methodology, theory, and intervention. Specifically, a method promotes methodological rigor through proper model specification, improved model fit, increased variance explained, strengthened causal inference, and greater clarification of the "real-world" structure of relations between variables (Mulaik, 2009). A method furthers theory building by enhancing the description, explanation, and prediction of ways in which stress-related disorders originate, persist, abate, recur, lead to dysfunction, and are counteracted by promotive and protective factors within and across specific developmental periods. Last, a method enhances intervention by improving the accuracy of risk detection and triage, identifying appropriate treatment components and modalities (e.g., classroom vs. small group treatment; Layne et al., 2008), and prioritizing intervention foci, such as targeting family versus peer versus adult mentor social support in a postwar psychosocial program (Layne, Warren, et al., 2009).

Some Essential Details: The Common Factor–Effect Indicator and Composite–Causal Indicator Models

Developing a method for accurately predicting who is at risk for what outcomes will involve evaluating two fundamentally different approaches to conceptualizing, measuring, and modeling trauma exposure. These are the *common latent factor with effect indicators model* (i.e., reflective–indicator model; MacDonald, 1999), which emphasizes the *structure of co-occurrence among events*, and the *composite variable with causal indicators model* (i.e., formative-indicator model; Grace & Bollen, 2008), which emphasizes the *structure of event occurrences and their respective consequences* (e.g., personal income and education both contribute to socioeconomic status). MacKenzie, Podsakoff, and Burke (2005) propose that the two models can be differentiated in four ways: (a) Effect indicators *manifest or reflect* the underlying construct, whereas causal indicators *define* the composite. (b) Effect indicators (assuming equivalent reliability) are *interchangeable*, whereas each (nonredundant) causal indicator is a *unique constitutive element of* the composite. (c) *Covariation* among effect indicators is both expectable and desirable due to the influence of a shared common factor, whereas causal indicators are not necessarily expected to covary (indeed, high covariation may reflect excessive redundancy). And (d) effect indicators, but not causal indicators, are expected to *share a common set of causal precursors and causal consequences* because they emanate from a common causal origin. The models can also be rationally differentiated by considering whether a change in the *common factor* necessarily precedes a change in its "effect" indicators (e.g., a siege *causes* a city to be surrounded and shelled), or conversely, whether a change in the "causal" indicators necessarily precedes a change in the resulting *composite* (surrounding and shelling a city *causes or creates* a siege).

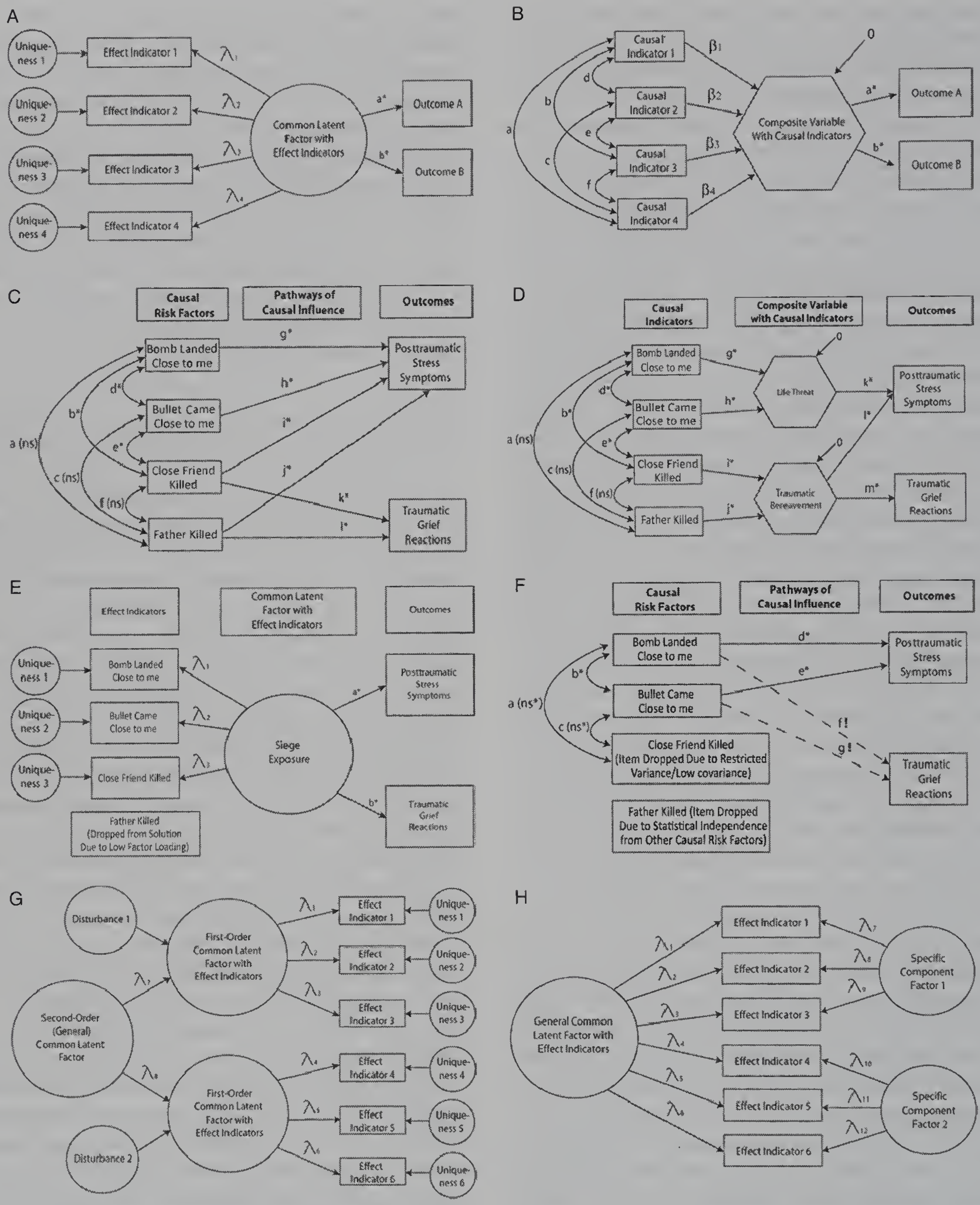

A
Uniqueness 1
Uniqueness 2
Uniqueness 3
Uniqueness 4
Effect Indicator 1
Effect Indicator 2
Effect Indicator 3
Effect Indicator 4
λ_1
λ_2
λ_3
λ_4
Common Latent Factor with Effect Indicators
a*
b*
Outcome A
Outcome B
B
Causal Indicator 1
Causal Indicator 2
Causal Indicator 3
Causal Indicator 4
a
b
c
d
e
f
β_1
β_2
β_3
β_4
0
Composite Variable With Causal Indicators
a*
b*
Outcome A
Outcome B
C
Causal Risk Factors
Pathways of Causal Influence
Outcomes
Bomb Landed Close to me
Bullet Came Close to me
Close Friend Killed
Father Killed
a (ns)
b*
c (ns)
d*
e*
f (ns)
g*
h*
i*
j*
k*
l*
Posttraumatic Stress Symptoms
Traumatic Grief Reactions
D
Causal Indicators
Composite Variable with Causal Indicators
Outcomes
Bomb Landed Close to me
Bullet Came Close to me
Close Friend Killed
Father Killed
a (ns)
b*
c (ns)
d*
e*
f (ns)
g*
h*
i*
j*
0
0
Life Threat
Traumatic Bereavement
k*
l*
m*
Posttraumatic Stress Symptoms
Traumatic Grief Reactions
E
Effect Indicators
Common Latent Factor with Effect Indicators
Outcomes
Uniqueness 1
Uniqueness 2
Uniqueness 3
Bomb Landed Close to me
Bullet Came Close to me
Close Friend Killed
λ_1
λ_2
λ_3
Siege Exposure
a*
b*
Posttraumatic Stress Symptoms
Traumatic Grief Reactions
Father Killed (Dropped from Solution Due to Low Factor Loading)
F
Causal Risk Factors
Pathways of Causal Influence
Outcomes
Bomb Landed Close to me
Bullet Came Close to me
Close Friend Killed (Item Dropped Due to Restricted Variance/Low covariance)
Father Killed (Item Dropped Due to Statistical Independence from Other Causal Risk Factors)
a (ns*)
b*
c (ns*)
d*
e*
f!
g!
Posttraumatic Stress Symptoms
Traumatic Grief Reactions
G
Disturbance 1
Disturbance 2
Second-Order (General) Common Latent Factor
λ_7
λ_8
First-Order Common Latent Factor with Effect Indicators
First-Order Common Latent Factor with Effect Indicators
λ_1
λ_2
λ_3
λ_4
λ_5
λ_6
Effect Indicator 1
Effect Indicator 2
Effect Indicator 3
Effect Indicator 4
Effect Indicator 5
Effect Indicator 6
Uniqueness 1
Uniqueness 2
Uniqueness 3
Uniqueness 4
Uniqueness 5
Uniqueness 6
H
General Common Latent Factor with Effect Indicators
λ_1
λ_2
λ_3
λ_4
λ_5
λ_6
Effect Indicator 1
Effect Indicator 2
Effect Indicator 3
Effect Indicator 4
Effect Indicator 5
Effect Indicator 6
λ_7
λ_8
λ_9
λ_{10}
λ_{11}
λ_{12}
Specific Component Factor 1
Specific Component Factor 2

Figure 1. (A) A common latent factor–effect indicators model. (B) A composite–causal indicators model. (C) Differentiated "real-world" causal structure between four causal risk factors and two outcomes. (D) A composite–causal indicators model of the variables in (C). (E) A common latent factor–effect indicators model of the variables in (C). (F) Incorrect causal inference regarding the variables in (C) produced by statistical artifact: Dropping *friend killed* forces both life threat indicators to act as its proxies. (G) A hierarchical confirmatory common latent factor–effect indicators model. (H) A bifactor confirmatory common latent factor–effect indicators model.

The *common factor–effect indicator* model (Figure 1A) derives from classical test theory and undergirds many traditional test construction tools, including exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and Cronbach's alpha (Cronbach, 1951; see also Bollen & Lennox, 1991). The common factor model, often paired with alpha, has been in use for over 20 years to derive "internally consistent" dimensions of trauma exposure (e.g., Dyregrov et al., 2000). A core assumption of the model is that the indicators are *homogeneous* (i.e., share a *common causal origin*; Borsboom, Mellenbergh, & van Heerden, 2003). Arrows leading from the latent common factor to each "effect" indicator place the locus of causation within the factor, signifying that it explains the variances of the indicators. The proportion of each indicator's variance explained by the common factor variance is denoted by its squared factor loading (λ^2). The model-implied covariance between any two indicators is reproduced by multiplying their factor loadings ($COV_{jk} = \lambda_j\lambda_k$; MacDonald, 1999), signifying that *indicators covary because, and to the extent that, they are jointly caused by causal processes underlying the common factor*. *U* denotes each item's uniqueness as comprised of both unsystematic (i.e., random) error variance and systematic variance not shared with the common factor (e.g., method variance). Although alpha assumes both item homogeneity (a strict assumption) and tau equivalence (equivalent factor loadings, a stricter assumption), it tests neither assumption but is instead a simple function of the number of test items and the average interitem bivariate correlation (Sijtsma, 2009).

In contrast, the *composite–causal indicator model* (Figure 1B; Grace & Bollen, 2008) does not assume indicator homogeneity. Rather, each of the indicators contributes to the composite (denoted by arrows leading from the indicators to the composite), placing the locus of causation in the indicators. Because each indicator is an independent contributor to the composite, covariation or "internal consistency" among indicators (Figure 1B, paths a–f) is of peripheral interest (Diamantopoulos, Riefler, & Roth, 2008). The error variance is specified at 0, signifying that composite values are a linear combination of the indicators. The composite thus mediates the predictive effects of its indicators on outcomes and constrains each indicator to have the same proportional influence on all outcomes (e.g., Indicator 1 exerts equivalent predictive effects on Outcomes A and B in Figure 1B; Franke, Preacher, & Rigdon, 2008). The model is flexible: Causal risk factors that are equifinal in relation to specified outcomes but known a priori to differ in their respective potencies can be aggregated into *fixed weight composites* (in which weights are fixed to different magnitudes to model differential effects); alternatively, risk factors of unknown potencies can be aggregated into *unknown weight composites*. Correctly specified composites function as "unitary entities" in relation to other variables (e.g., causal indicators theorized to be equifinal should relate in similar ways to their theorized causal consequences; Wilcox, Howell, & Breivik, 2008). Accordingly, equifinal causal risk factors (e.g., father's death by shelling, aunt's death by suicide, brother's death by car accident) may exert similar effects on a specified set of outcomes (PTSD, traumatic grief) *irrespective* of whether they share a common origin, co-occur, covary, or manifest "internal consistency." Thus, *when the aim is to model links between event occurrences and their respective consequences*, risk factors that *are* equifinal in reference to focal outcomes may be assigned to the same composite, whereas risk factors that co-occur, covary, and manifest internal consistency, but are *not* equifinal, may be assigned to different composites. Given its focus on modeling consequences, the composite approach is well suited for modeling not only relations between "real-world" objective events (e.g., paths from war-time losses to postwar hardships) but also between objective events and subjective experiences (e.g., paths between trauma and distress). The degree of discrimination used in combining events into dimensions depends on the desired degree of specificity in measuring their differential effects.

Contrasting the Two Approaches: Conceptualizing, Measuring, and Modeling Trauma Exposure

A comparison of the common factor versus composite models should be qualified by two caveats. First, neither is entirely theory driven nor entirely data driven: Although strongly empirical, the common factor approach requires rational judgment pertaining to such activities as selecting extraction and rotation algorithms, specifying factor-loading cutoff values, and factor interpretation. Further, the composite approach should be guided by both theory and empirical evidence concerning links between causal risk factors and their consequences. Thus, the approaches may be conceptualized as lying on opposing ends of a bipolar continuum ranging from greater emphasis on empirical evidence to greater emphasis on theory.

A second caveat is that the strengths and weaknesses of each model depend on the application. Consider the "real-world" differentiated model in Figure 1C, in which all risk factors are equifinal for PTSD symptoms, but only *father killed* and *close friend killed* are equifinal for traumatic grief. The utility of each approach rests on which portion of the model—the *structure of relations between co-occurring risk factors*, versus the *structure of the occurrence of specific causal risk factors and their differential consequences*—is of primary interest. Because the common factor model draws exclusively on the *covariances between risk factor indicators* (paths a–f), it is useful for aggregating clusters of traumatic events that co-occurred sufficiently to produce coherent factors. This focus on the structure *among* risk factors gives the common factor model three potential strengths. First, each indicator is a risk marker for the other indicators. Second, all indicators are *general* risk markers for all outcomes (APA Presidential Task Force, 2008). Third, the model can ''unpack'' complex events (e.g., wars) into often-meaningful clusters of co-occurring events (e.g., sieges). However, because it is designed to extract ''best internally consistent'' *samples* of covarying events, the model may drop events that *occur, but do not systematically co-occur*, that would be captured via event census (Netland, 2005). Indeed, the only conditions under which the common factor model is likely to produce ''clean'' equifinal test dimensions are when all equifinal events (e.g., all types of traumatic bereavement) happen to covary strongly, and all *non*equifinal events (bereavement vs. material loss) do not. Such conditions are unlikely to emerge in war-related settings, although there may be rare exceptions (cf. Macksoud & Aber, 1996).

Figure 1D presents a composite model of the relations in Figure 1C. Given its emphasis on causes and consequences, including differential relations (see paths g–l), this approach conducts a *census* of trauma exposure types for two reasons. First, each (nonredundant) causal indicator is a constitutive element of the composite and thus indispensable (Bollen & Lennox, 1991). Second, events that occur but do not *co-occur and covary* with other risk factors may nevertheless exert potent influences (see Figure 1C, paths j and l from *father killed*, which does not covary). The ''data input'' used by the composite model thus centers on *paths linking risk factors to their theorized consequences* (Figure 1C, paths g–l). The composite model is thus well suited not only to aggregate equifinal risk factors into common dimensions, but also to assign causally diverse events to different composites to facilitate the detection of differential relations.

Model–Application Mismatch: Using the Common Factor Model to Unpack Risk Factor–Consequence Links

Given its long-standing use (see Barenbaum et al., 2004; La Greca, Silverman, Vernberg, & Roberts, 2002), potential drawbacks that may arise when the common factor model is used *in applications focusing on the causal consequences of war exposure* should be considered. A model that ''packs'' types of war exposure into dimensions based on their covariance structure (Figure 1C, paths a–f) may indeed be ill suited to ''unpack'' the differential causal effects *of* specific types of exposure on specific consequences (paths g–l). In such applications, any of six conditions may create ''model–application mismatches'' that dilute, distort, and obscure ''real-world'' differential relations. These include: (a) The aim is to detect differential relations between specific types of causal risk factors and specific outcomes. (b) Events that co-occur are *not* equifinal (Figure 1C, paths g–j). (c) Events that do not co-occur *are* equifinal (paths g, h, and *k*). (d) Ambiguity in test scores (produced by combining different causal risk factors that exert different effects) is costly. (e) *Sampling*, rather than conducting a *census* of, trauma types will create costly false negatives (e.g., youths whose fathers are killed are not detected), costly false positives (youths needing only general support receive specialized services), or lead to incorrect conclusions about causal relations. And (f) researchers wish to create variables and assessment tools that have predictive utility at the individual case level and are transportable across diverse settings and diverse traumatic events (APA Presidential Task Force, 2008).

Figure 1E depicts a scenario that may arise when the common factor model is used to unpack the ''real-world'' consequences of war exposure (as found in Figure 1C) and contains three major misspecification errors. These include the *violation of the assumption of homogeneity*, the *loss of differential relations*, and the *exclusion of an influential causal risk factor*. The first error, *violation of indicator homogeneity*, arises because *Siege Exposure* is modeled as the *cause* of shooting and shelling, when instead *these acts define and initiate a siege and ceasing them will cause it to end* (MacKenzie et al., 2005). This fact repudiates the assumption of indicator homogeneity and the associated need to aggregate (or disaggregate) indicator-level risk factors into test dimensions on the basis of their covariance structure. Why should causal risk factors (including those that exert equifinal effects) necessarily co-occur or covary if they do not share a common causal origin? Indeed, modeling war exposure as a homogeneous latent common factor is incompatible with PTSD diagnostic criteria because the etiology of the disorder lies in observable *indicator-level* Criterion A1 events, such as mortar shell attacks, rather than unobservable ''latent'' causal processes. The indicators thus do not *share* a common origin with outcomes (e.g. ''latent'' sieges do not cause shelling on one hand *and* PTSD on the other); rather, *indicators are etiological risk factors for PTSD* (*shelling* can cause PTSD). Causally misspecified common factor models may thus lead to the logical fallacy of *affirming the consequent*: (a) If test items are homogeneous, then they load a common factor and show evidence of internal consistency. (b) This set of indicators loads a common factor and produces a high alpha value. (c) *Therefore, these indicators are homogeneous and share the same causal consequences* (MacKenzie et al., 2005). Thus, well-intentioned researchers may create a test of "siege exposure" and attempt to model, assess, or treat its "sequelae."

The second specification error found in Figure 1E is the *loss of differentiated relations* (see Figure 1C, paths k and l). Sieges and other campaigns often generate clusters of co-occurring events (Figure 1C, paths b, d, and e) that a common factor approach will aggregate. However, covariation between types of war exposure does not reflect indicator homogeneity, but rather, *perceived event co-occurrences*. "Packing" types of war exposure into dimensions on the basis of their covariance structure is thus likely to intermix causally diverse (i.e., nonequifinal) risk factors into the same dimension. Such causally diverse dimensions are especially likely to emerge in inherently complex war and disaster settings where many different types of trauma and loss co-occur. Causally diverse test dimensions are of concern for two reasons. First, creating summative ("unit-weighted") aggregate variables to model or assess the "sequelae" of a given dimension of war exposure implies that *all indicators are equifinal and equally potent* causal risk factors for all its modeled sequelae. For example, summative scoring of the model in Figure 1E implies that both life-threat events and *friend killed* are *equivalent and interchangeable* causal risk factors for both PTSD and traumatic grief (MacKenzie et al., 2005). Second, pooling the variances of causally diverse risk factors into the common factor variance (thereby increasing *intradimension* causal diversity) decreases their sensitivity to detecting the differential effects of different types of risk factors. Indeed, major discrepancies may arise between the *causal effects* of risk factors in the "real world" and the *magnitudes of their factor loadings*. In Figure 1C, the effects of the life-threat items on traumatic grief are negligible, whereas *friend killed* and *father killed* may exert moderate and potent effects, respectively. However, the covariances among indicators (paths a–f) may produce large loadings for both life-threat items and moderate and negligible loadings for *friend killed* and *father killed*, respectively. The common factor variance may thus consist mostly of variance extracted from causally weak indicators. This may both reduce the strength of paths from the common factor to outcomes (Figure 1E, path b) and lead to erroneous conclusions regarding the potencies of the respective indicators or of that dimension.

The third misspecification error in Figure 1E is the *loss of a potent causal risk factor* (*father killed*) due to its lack of covariation with other risk factors (Figure 1C, paths a, c, and f). Because it extracts variables based on their covariance structure, the common factor model is susceptible to dropping two types of events. These include *events that occur but do not systematically co-occur* with other modeled risk factors (e.g., random or unrelated events), and *events whose measurement is compromised by statistical artifact* (e.g., restricted variance due to floor effects). The exclusion of causal risk factors may reduce proportions of explained variance (reducing predictive precision) and bias model parameters in ways that overestimate the vulnerability of exposed subgroups (due to "unexplained" elevations in distress scores; Netland, 2005). Omitting causal risk factors also compromises test validity by reducing the correspondence between a test and the "real-world" structure among variables it purports to measure (Borsboom, Mellenbergh, & van Heerden, 2004) and can decrease the sensitivity of screening instruments in ways that lead to false negatives and poor matches of individuals to treatments (Kraemer et al., 1999; Netland, 2001). Omitted causal risk factors can also create spurious statistical artifacts. Figure 1F depicts how distortions of the "real-world" scenario in Figure 1C can lead to faulty conclusions about causal relations. Statistically independent *father killed* does not covary, whereas high endorsement rates for *friend killed* produce a ceiling effect that artificially restricts its covariance with both life-threat items *even though these three types of events co-occur in that setting*. The exclusion of *friend killed* forces both life-threat items to act as its proxies. The resulting biased model parameters may undermine theory building by *under*estimating the etiological significance of traumatic death in relation to traumatic grief, while *over*estimating that of life threat (Netland, 2005). The model may also impede intervention by increasing false negatives ("high-risk" bereaved youths are not detected) and false positives (youths exposed to low-potency risk factors receive specialized treatment components).

Appropriate Use of the Common Factor–Effect Indicator Model to Clarify Differential Relations

Notwithstanding these caveats, we emphasize the utility of the common factor model for clarifying differential relations among multidimensional constructs whose causal structure is compatible with indicator homogeneity (e.g., Krueger, 1999). These include both the *hierarchical* and the *bifactor* confirmatory factor analytic models (Figures 1G and 1H; Chen, West, & Sousa, 2006). Both models allow a priori model specification of hypothesized dimensions and carry a lower risk of bias in model parameters than EFA (Little, Lindenberger, & Nesselroade, 1999). The latent structure in both models consists of a general factor and one or more specific components for different subsets of items. Although "specific" components are often used to model potentially confounding methodological effects, they can also model multidimensional structures among "real-world" variables. In contrast to traditional common factor models that contain multiple *separate* factors, the hierarchical and bifactor models each allows an item to reflect both general characteristics shared with all of the other indicators and characteristics unique to specific item subsets. The design of the bifactor model ensures that the latent-specific components are uncorrelated with one another and with the general factor. These features make the models useful for studying the breadth and specificity of complex constructs, including *convergent validity* (via the general factor) and *discriminant validity* (via the specific components) in multifaceted conceptual domains. The models permit testing of measurement and structural hypotheses relating to differential relations while addressing potential collinearity among latent components and controlling for broader shared variance. Although the hierarchical and bifactor models are difficult to distinguish on purely empirical grounds, each embodies a distinctive theoretical structure that may be plausible within a given setting. Moreover, the common factor approach can also, like the composite–causal indicator approach, prioritize the clarification of structural relations between constructs over internal consistency. Little et al. (1999) demonstrate that the validity of common factor models often increases when indicators are selected (even those of "poor" psychometric quality) that clearly demarcate construct domains and boundaries between constructs. Conversely, selecting only highly correlated indicators in the interests of maximizing internal consistency may lead to test bias and distorted estimates of the construct's relations with other constructs.

A Comparison of the Common Factor and Composite-Based Approaches in Postwar Bosnian Youth

Study rationale. Given the potential benefits of each approach, as well as potential problems arising from model–application mismatches, this study compared the performance of a common factor–effect indicator model (using EFA) versus a composite–causal indicator model. The study focused on the potential of each model to (a) build theory capable of explaining differential risk factor–outcome relations, (b) predict which groups are at risk for specific types of distress, and (c) increase the effectiveness and efficiency of interventions. Because low endorsement rates restricted the variances of many war exposure items (e.g., personal injury), the composite approach relied only on theory and prior knowledge to assign specific types of trauma to exposure dimensions. Further, because the cross-sectional design precluded causal inference, we investigated only the *predictive effects* of risk factors (including theorized mediators) on outcomes. Last, given space limitations and our focus on "unpacking" the covariance structure and consequences of risk factors, no gender effects were tested.

Study setting. The 1991–1995 civil war in the Former Yugoslavia exemplifies the brutal nature of modern armed conflict and its stressful aftermath. The 1992–1995 Bosnian conflict led to the traumatic deaths of close to 100,000 people, approximately 41% of whom were civilians, and the displacement of 1,800,000 people, many of whom faced years of subsequent hardship as internally displaced persons or war refugees (Ahmetasevic, 2007). The war led to widespread traumatic deaths, physical injury, traumatic disappearances, major disruptions in social institutions, and brutal mistreatment through acts of assault, torture, rape, and starvation inflicted on detained civilians and prisoners of war. Many military strategies directly targeted civilians, including sieges of cities and protected enclaves, ethnic cleansing campaigns, and genocidal operations. These campaigns generated widespread exposure to direct life threat, the witnessing of traumatic deaths and serious injury, deprivations in basic necessities, exposure to the elements, and the forced exodus of entire regions in refugee flights (Powell & Duraković-Belko, 2002).

Study design and hypotheses. Data were collected in the fall of 2000 (nearly 5 years postwar) in a screening survey used in a postwar psychosocial program for war-exposed adolescents (Layne et al., 2008). The survey was administered in nine heavily war-exposed secondary schools in Central Bosnia and was based on a developmental model of childhood traumatic stress (Pynoos, Steinberg, & Wraith, 1995). Survey items measured a broad range of traumatic experiences, theorized mediators of youth postwar adjustment including trauma reminders and adversities (Layne et al., 2006), and distress-related outcomes. Because the common factor approach models the *structure of event co-occurrences*, we predicted that EFA-derived factors would reflect major military strategies employed in the Bosnian conflict (e.g., sieges, ethnic cleansing; Hypothesis 1) and would substantively differ in their composition compared with the rationally derived composite dimensions (Hypothesis 2). We also theorized that highly "traumatogenic" war-time risk factors (e.g., life threat, traumatic death) would exert direct predictive effects, whereas the predictive effects of less-traumatogenic war-time risk factors would be fully mediated by postwar trauma reminders and adversities. We thus predicted that the *partially mediated* version of each model would fit better than its *fully mediated* (indirect effects only) counterpart (Hypothesis 3). Last, we predicted that the composite–causal indicator model would show more differentiated relations between theorized causal risk factors (including *prewar trauma* and different types of *war-time trauma*), theorized mediators (including *postwar adversities* and *trauma reminders*), and outcomes (including *posttraumatic stress, depressive, and traumatic grief reactions*; and *functional impairment*) than the common factor model (Hypothesis 4).

Method

Measures

All measures were reviewed for cultural appropriateness and forward- and back-translated by doctoral students in psychology at the University of Sarajevo. Measures were then combined into a self-report screening survey and administered to classrooms by trained school counselors. All distress measures except the OQ Somatization Subscale used a standard 5-point Likert-type scale measuring the frequency with which each symptom was experienced during the last month ranging from 0 (*none of the time*) to 4 (*most of the time*). All trauma and stressful life event exposure measures used a dichotomous scale ranging from 0 (*event did not occur within the specified time frame*) to 1 (*event occurred at least once within the time frame*). Prior to modeling, we transformed the observed score on each scale or subscale by dividing by the total number of items to form an *item-level average score* ($\bar{x}_i$). Calculating the item-level average enhanced interpretation by benchmarking the summative scale score against the original item-level metric (averages of Likert scale scores ranged from 0 to 4; averages of exposure scale scores ranged from 0 to 1) while controlling for between-scale differences in scale length and scale variance.

Participants

Participants consisted of ($N = 881$) students attending nine Central Bosnian secondary schools during the 2000–2001 school year who were screened for their appropriateness for a UNICEF-sponsored trauma and grief-focused psychosocial program (Layne et al., 2008). The sample consisted of 55% ($n = 480$) girls and 44% ($n = 390$) boys (1% not reported), ranging from 13 to 19 years old ($\overline{X} = 15.79$, $SD = 1.10$) and from the 1st to the 3rd year in school ($\overline{X} = 1.87$, $SD = 0.82$). We randomly assigned approximately half of the participants to one of two subsamples. We used Subsample 1 for model construction, and Subsample 2 to replicate (without further modification) the model parameters constructed using Subsample 1. The two subsamples did not significantly differ ($p < .05$) in age, sex, grade in school, or school affiliation.

Measures

Prewar Trauma Exposure Index (Layne, 1999a) is a 10-item self-report measure of exposure to a range of prewar traumatic experiences. Items were generated from field research and from adapting items from the Stressful Life Events Screening Questionnaire (SLESQ; Goodman, Corcoran, Turner, Yuan, & Green, 1998). Items are measured on a yes–no scale ("Before the war, did you ever have a life-threatening illness?"). Items 1–9 (the 10th is open-ended) were summed into a composite score (total sample $\overline{x_i} = 0.08$).

War Trauma Screening Inventory (WTSI; Layne, Stuvland, Saltzman, Djapo, & Pynoos, 1999) is a 49-item self-report measure of exposure to different types of war-related trauma and loss. Items were drawn from Bosnian field research and from a literature review (e.g., Macksoud & Aber, 1996). Items were analyzed in two ways: For the composite–causal indicator approach, the first author drew on proposed "common denominator" dimensions of trauma exposure (e.g., La Greca et al., 2002; Pynoos

Table 1
Composite–Causal Indicator-Derived Dimensions of War Trauma Exposure

Direct Exposure	Witnessing Violence	Life Threat	Traumatic Death[a]	Harm to Loved Ones	Threat to Loved Ones	Loss and Displacement	Separation from Loved Ones
In serious accident Life-threatening cold Deprived of food/water Exposed to disaster Physically tortured Taken prisoner Unwanted sexual contact Was physically injured Victim of serious crime Became seriously ill Was physically assaulted	Saw dead body Saw severely injured person Saw serious accident Saw property destruction Saw someone being killed or injured Touched/carried wounded or killed person Saw distraught person Witnessed assault Witnessed person taken prisoner or abducted Witnessed torture Witnessed loved one being harmed or killed	Grenade/bomb exploded nearby Bullet came close Strongly believed would be hurt or killed Verbally threatened with harm/death Enemy soldiers entered home	Close friend killed Mother killed Sibling killed Extended family member killed Loved one committed suicide Loved one killed in violence not due to war Father killed	Loved one tortured Loved one assaulted Loved one taken prisoner or detained Loved one seriously injured Loved one had life-threatening illness	Loved one had very dangerous duties Loved one threatened with harm/death Loved one went missing	Left country due to war Forced from home Left village/town because of war Changed to new school because of war Lived in refugee center Home badly damaged Lost beloved pet	Separated from loved one when greatly feared for their safety Separated from both caregivers for month or longer

[a] The item *loved one died of natural causes* was dropped from the composite–causal indicator-derived dimensions due to its ambiguity as a traumatic stressor.

et al., 1995) to divide items into eight content domains consisting of *Direct Exposure* (11 items; total sample $\overline{x_i} = 0.066$), *Witnessing Violence* (11 items, $\overline{x_i} = 0.307$), *Life Threat* (5 items, $\overline{x_i} = 0.354$), *Traumatic Death* (7 items, $\overline{x_i} = 0.055$), *Harm to Loved Ones* (5 items, $\overline{x_i} = 0.199$), *Threat to Loved Ones* (3 items, $\overline{x_i} = 0.309$), *Loss and Displacement* (7 items, $\overline{x_i} = 0.341$), and *Separation From Loved Ones* (2 items, $\overline{x_i} = 0.362$). Table 1 presents a breakdown of WTSI items by the eight domains. For the common factor–effect indicator method, five war exposure subscales were derived using EFA (see Table 2) consisting of *Siege Exposure* (18 items including 1 reverse-scored item, total sample $\overline{x_i} = 0.369$), *Expulsion and Displacement* (9 items including 2 reverse-scored items, $\overline{x_i} = 0.459$), *Detention and Maltreatment* (9 items, $\overline{x_i} = 0.095$), *Harm to or Death of Loved One* (7 items, $\overline{x_i} = 0.228$), and *Personal Assault or Injury* (2 items, $\overline{x_i} = 0.03$).

Postwar Trauma Exposure Inventory (PWTEI; Layne, 1999b) is a 13-item self-report measure of exposure to nonwar-related traumatic events. Items were adapted from the SLESQ and from field research in postwar Bosnia. Items include general traumatic events ("Since the war, has a loved one attempted suicide?") as well as events characteristic of the Bosnian conflict ("Since the war, has a loved one who disappeared during the war been declared dead?"). The PWTEI contains four content domains that were rationally derived by the first author, including *Direct Exposure* (6 items, total sample $\overline{x_i} = 0.062$), *Witnessing* (2 items, $\overline{x_i} = 0.123$), *Harm to Loved Ones* (3 items, $\overline{x_i} = 0.101$), and *Traumatic Death* (2 items, $\overline{x_i} = 0.072$).

Trauma Reminders Screening Inventory (TRSI; Layne, Steinberg, & Pynoos, 1998) is a 17-item self-report measure of the frequency of respondents' exposure to reminders of war-related traumatic events during the past month and associated functional impairment. Using theory and EFA, items were partitioned into *general trauma reminders* (8 items; e.g., news of war atrocities, political instability), *interpersonal trauma reminders* (4 items; e.g., friends, family), and *trauma reminder-related interference* (3 items; e.g., impaired social relationships). Because the TRSI measures *perceptions* of being reminded (suggesting a common underlying causal process), alphas were calculated for each subscale (full-sample general α = .90, interpersonal α = .83, interference α = .85). We modeled general and interpersonal reminders as postwar mediating variables, and reminder-related interference as an outcome.

UCLA Reaction Index–Revised (RI–R; Steinberg, Brymer, Decker, & Pynoos, 2004) as used in this study is a 17-item self-report scale of symptom frequency during the previous month. Items correspond to the 17 *DSM–IV* PTSD criteria. The total scale has shown good internal consistency (α = .87), criterion-referenced validity in relation to a range of distress measures (.30 to .70), and 2-week test–retest reliability (.75) in postwar Bosnian youth (Layne, Warren, et al., 2009). Full sample, full-scale α = .87.

Depression Self-Rating Scale (Birelson, 1981) is an 18-item self-report measure of depressive symptoms. The original 3-point scale (for children) was modified to a 5-point frequency scale to increase developmental appropriateness and sensitivity to clinical change. The adapted version has shown acceptable internal consistency (α = .85), criterion-referenced validity (.37 to .62), and test–retest reliability (.64) in postwar Bosnian youth (Layne, Warren, et al., 2009). Full sample, full-scale α = .89 in this study.

Youth Outcome Questionnaire: Somatic Symptoms Subscale (Wells, Burlingame, Lambert, Hoag, & Hope, 1996) is an 8-item self-report measure of somatic distress (e.g., "I have headaches or feel dizzy") in children and adolescents. Items are measured on a 5-point frequency scale ranging from 0 (*never or almost never*) to 4 (*almost always or always*). Full sample, full-scale alpha α = .80 in this study.

Postwar Adversities Index (Layne & Djapo, 1999) is a 25-item self-report inventory of stressful life events and circumstances during the past 6 months. Items were rationally divided into two content domains by two independent raters who showed substantial interrater agreement using a dichotomous scale (κ = .79). Domains were *postwar existential adversities* (16 items; short on pocket money) and *interpersonal adversities* (6 items; living with people who fight). Items were coded according to valence (positive/neutral/negative life event) and independence/dependence by six independent raters on a 5-point scale. Raters showed excellent interrater reliability (α = .92) on valence and acceptable reliability (α = .79) on event independence. Chronic and acute events (Masten, Neemann, & Andenas, 1994) were both included. Because this study focused on adverse outcomes, only *negative-independent* life events were included; three events rated as *dependent* (pregnancy, getting into fights, and drug use) were dropped from analysis.

Procedure

To minimize potential stigmatization of individual students, school counselors selected entire classrooms with the highest prevalence rates of

Table 2

Percent Item Endorsement (%), Factor Loadings (F_i), Communalities (h^2), and Factor Correlations for Maximum Likelihood Factors Extraction and Promax Rotation on War Exposure Inventory Items

Item (type of trauma exposure during war)	%	F_1[a]	F_2	F_3	F_4	F_5	h^2
Bomb landed so close could have been hurt/killed	56	**0.77**	−0.02	−0.14	−0.02	0.02	0.54
Saw the body of someone who had been killed	45	**0.76**	−0.23	0.12	−0.10	−0.00	0.58
Bullet came so close could have been hurt/killed	53	**0.75**	0.02	−0.07	−0.02	0.09	0.56
Saw someone who was severely injured	56	**0.73**	−0.12	0.11	0.04	−0.10	0.57
Witnessed a serious accident	20	**0.69**	−0.06	−0.03	0.05	0.24	0.59
Witnessed the massive destruction of property	51	**0.68**	0.05	−0.01	−0.04	0.02	0.45
Witnessed someone being killed/severely injured	21	**0.65**	−0.15	0.35	−0.01	−0.09	0.61
Was in serious accident (car crash, bad fall, mine)	05	**0.65**	−0.04	−0.16	0.03	0.17	0.44
A close personal friend was killed	34	**0.57**	0.04	−0.15	0.19	−0.03	0.42
Touched/carried seriously injured or killed person	11	**0.55**	−0.18	0.13	0.01	0.22	0.44
Experienced life-threatening cold	14	**0.53**	0.07	0.09	−0.04	−0.00	0.32
Witnessed people in extreme distress (distraught)	72	**0.51**	−0.16	0.11	0.38	−0.21	0.54
Life-threatening deprivations of food or water	13	**0.48**	0.12	0.07	0.01	0.13	0.34
Strongly believed would be seriously hurt or killed	51	**0.45**	0.05	0.17	0.23	−0.12	0.42
Exposed to major natural or man-made disaster	12	**0.43**	0.19	0.02	−0.08	0.25	0.32
Loved one had very dangerous duties	79	**0.41**	0.07	−0.29	0.32	−0.13	0.38
Witnessed someone being physically assaulted	20	**0.40**	0.13	0.32	0.03	−0.11	0.38
Left country because of the war	18	**−0.51**	**0.56**	0.05	0.03	0.43	0.62
Forced to leave home because of war	48	0.02	**0.91**	0.16	−0.03	−0.07	0.90
Forced to leave village/town because of war	37	−0.28	**0.90**	0.07	0.06	0.08	0.86
Forced to change to new school because of the war	49	−0.01	**0.74**	0.09	0.07	0.17	0.64
Lived in a collective refugee center during the war	11	−0.15	**0.62**	0.04	0.10	0.25	0.49
Home seriously damaged in the war	35	0.26	**0.51**	0.07	−0.04	−0.04	0.38
Lost a beloved pet because of war	26	0.22	**0.47**	0.03	0.06	−0.07	0.34
Mother killed during war	01	−0.07	**−0.46**	0.35	0.23	0.30	0.42
Brother or sister killed during war	34	0.21	**−0.43**	−0.02	0.17	0.27	0.29
Was physically tortured	02	−0.04	−0.01	**0.76**	−0.01	0.38	0.86
Taken prisoner or confined in detention camp	03	−0.14	0.11	**0.73**	0.09	0.23	0.71
Was verbally threatened with harm or death	05	−0.02	−0.08	**0.70**	0.13	0.09	0.55
Witnessed someone being abducted/taken prisoner	22	0.25	0.16	**0.70**	−0.06	−0.32	0.66
Enemy soldiers forcibly entered residence	12	−0.06	0.30	**0.66**	0.15	−0.08	0.65
Witnessed someone being tortured	08	0.38	0.13	**0.55**	−0.20	−0.06	0.53
Loved one verbally threatened with harm or death	11	0.03	0.20	**0.53**	0.40	−0.26	0.64
Witnessed a loved one being harmed or killed	09	0.20	−0.08	**0.50**	0.19	0.23	0.56
Loved one was physically assaulted	14	0.03	0.11	**0.49**	0.44	−0.25	0.56
Loved one was tortured	08	−0.09	−0.05	0.49	**0.54**	0.09	0.60
Loved one missing, greatly feared for their safety	20	−0.02	0.12	0.18	**0.59**	0.05	0.49
Loved one taken prisoner/held in detention camp	20	−0.14	0.10	0.35	**0.51**	0.11	0.50
Extended family member killed	47	0.16	0.03	−0.17	**0.51**	−0.12	0.34
Loved one was seriously physically injured	37	0.19	−0.02	−0.09	**0.45**	0.19	0.34
Loved one committed suicide	05	−0.05	−0.09	0.04	**0.44**	0.12	0.20
Loved one died in accident or community violence	08	0.01	0.08	0.01	**0.43**	0.27	0.33
Was touched in a way that I did not want	02	0.04	0.09	0.06	0.07	**1.07**[b]	1.24
Seriously injured during war	04	0.28	−0.05	0.18	−0.07	**0.59**	0.55
***Nonloading items (all loadings < 0.40):**							
Father killed during war	10	−0.04	−0.16	0.04	0.33	−0.04	0.10
Separated from loved one when greatly feared for their safety	60	0.32	0.28	−0.14	0.25	0.06	0.34
Loved one died of natural causes	47	0.22	0.24	−0.22	0.16	−0.15	0.21
Loved one had life-threatening physical illness	14	0.27	0.05	−0.05	0.08	0.05	0.10
Was the victim of a serious crime	06	0.02	0.25	0.39	0.01	0.16	0.32
Separated from primary caregivers for ≥1 month	13	0.15	0.29	−0.02	−0.02	0.32	0.23
Became seriously ill	10	0.23	0.12	−0.06	−0.05	0.32	0.18
Was physically assaulted	03	0.34	−0.08	0.38	−0.06	0.26	0.44

Table 2
Continued

Item (type of trauma exposure during war)	%	F_1[a]	F_2	F_3	F_4	F_5	h^2
Promax factor correlations	F_1	1.00					
	F_2	.18	1.00				
	F_3	.27	.21	1.00			
	F_4	.14	.04	.27	1.00		
	F_5	.37	.31	.21	.13	1.00	

[a] Factor labels: F_1 = siege exposure; F_2 = expulsion and displacement; F_3 = detention and maltreatment; F_4 = harm to or death of loved one; F_5 = direct personal assault or injury. Primary factor loadings for each variable are bolded (decision rule: complexly loaded items with both loadings ≥ 0.50 load on both factors). [b] Heywood case (model estimates a negative residual variance) linked to low item variance (Comrey, 1992).

severely war-exposed students. Details regarding classroom selection, consent, and survey administration procedures are provided elsewhere (Layne et al., 2008). Measures and procedures were approved by an ad hoc institutional review board (IRB) formed of Bosnian mental health professionals and a UNCEF officer, and by the IRB of Brigham Young University.

Results

In a first step, we used Mplus 5.1 (Muthén & Muthén, 2009) to perform EFA on the full-sample covariance matrix formed by the 49 WTSI items. Maximum likelihood factors extraction was used in combination with promax (oblique) rotation. This procedure adjusts for missing values on observed variables and is suitable for binary-ordered categorical indicators (Muthén, 2002). After several preliminary extractions to estimate the number of factors and verify the absence of multicollinearity and the factorability of the correlation matrix, five factors were extracted. Table 2 presents item endorsement rates (in %), the rotated factor pattern matrix, communality (h^2) values, and a suggested label for interpreting each factor. Given the dichotomous scaling of the WTSI and low (< 10%) endorsement rates for many items, a liberal criterion ($\lambda_j \geq .40 = 16\%$ explained variance) was used to denote the primary factor loadings. The factors varied in quality as indicated by the number of variables with primary loadings and the sizes of the loadings. Factors were labeled *Siege Exposure* (Factor 1), *Expulsion & Displacement* (Factor 2), *Detention & Maltreatment* (Factor 3), *Harm to or Death of Loved One* (Factor 4), and *Direct Personal Assault or Injury* (Factor 5). Although interpretable, the solution poorly accounted for many items as denoted by low communalities and the loss of eight items (15% of total) due to low loadings. Among dropped items were events with low (*mother killed* = 1%), moderate (*father killed* = 10%), and comparatively high endorsement rates (*physically separated from loved one* = 60%) that correlated weakly with the retained items (full correlation matrix available from first author). Consistent with Hypothesis 1, the five EFA-derived factors reflected idiosyncratic features of the Bosnian war, including sieges, ethnic cleansing, and detention and maltreatment (Powell & Duraković-Belko, 2002).

In a second step, we evaluated the degree of convergence between the trauma exposure dimensions derived using the common factor versus the composite approach. The approaches were compared at two levels: (a) visually inspecting the item contents of the factors and composites (i.e., comparing Table 1 with Table 2), and (b) examining correlations between the respective subscale scores (a comparison matrix of factor/composite item overlaps and intercorrelations is available from the first author). In support of Hypothesis 2, the two approaches showed poor to moderate convergence at both levels of analysis and were consistent with our concerns regarding the *within-dimension* diversity of war exposure dimensions derived via the common factor model. Visual inspection revealed that *Siege Exposure* (Factor 1), *Detention & Maltreatment* (Factor 3), and *Harm to or Death of Loved One* (Factor 4) each contained a broad range of trauma types, such that their constituent items were broadly dispersed across the composite dimensions. For example, the item contents of Factor 1 (*Siege Exposure*) were distributed across the *Direct Exposure, Witnessing Violence, Life Threat, Traumatic Death, Threat to Loved Ones,* and *Loss and Displacement* composites. Further, the contents of Factor 3 (*Detention & Maltreatment*) were distributed across *Direct Exposure, Witnessing Violence, Life Threat,* and *Harm to Loved Ones*. The greatest degree of convergence was found in the overlap between *Expulsion*

& Displacement (Factor 2) and *Loss and Displacement* (Composite 7). The inverse loadings of *mother killed* and *sibling killed* on *Expulsion & Displacement* were unexpected. A second point of convergence was found in the overlap between the two items loading *Direct Personal Assault or Injury* (Factor 5) and *Direct Exposure* (Composite 1). However, because *Direct Exposure* contained six items not found in Factor 5, Factor 5 was a small subset of Composite 1.

The general lack of convergence between the two approaches observed in our visual inspection was reflected in a similar lack of differentiation in their subscale intercorrelations. The strongest points of convergence were found between *Siege Exposure* (Factor 1) and *Witnessing Violence* (Composite 2) ($r = .89$), and between *Expulsion & Displacement* (Factor 2) and *Loss and Displacement* (Composite 7) ($r = .97$). More generally, however, the factor–composite correlation matrix showed little clear evidence of between-method convergence between specific factors and composites, with the great majority of correlations falling in the low to moderate range ($.10 < r < .60$). For example, the Factor 3 (*Detention & Maltreatment*) subscale score correlated between $r = .40$ and $.60$ ($p < .01$) with those of the *Direct Exposure*, *Witnessing Violence*, *Life Threat*, and *Harm to Loved Ones* composites. In general, most factor subscale scores correlated indiscriminately with most composite subscale scores—an unsurprising finding given the high degree of dispersion of factor item contents across composites, and vice versa.

In a third step, we used Amos 16.0 (SPSS, 2009) to evaluate the fit of both the common factor and composite models in Subsample 1, and to evaluate the fit of the (unmodified) models in Subsample 2. We formed our models using observed variable path analysis—a variant of SEM conducive to modeling composite variables (Byrne, 2001). Given the exploratory nature of the study and the lack of guiding theory regarding relations between *specific* types of trauma and loss exposure, mediators, and outcomes, we used a conservative ''repeated pruning'' model-building strategy for both the composite and common factor method subsamples: We first created a semisaturated model that contained all possible paths linking (a) prewar trauma to each theorized mediator and each outcome, (b) each war-time exposure dimension to each theorized mediator and each outcome, and (c) each theorized mediator to each outcome. We then tested the model and deleted all nonsignificant paths with $p > .10$. We then retested the model and deleted all remaining nonsignificant paths with $p > .10$. In a final pass, we retested the model and deleted all nonsignificant paths with $p > .05$.

The composite–causal indicator model for Subsample 1 (see Figure 2) had good fit, $\chi^2(79) = 75.73$, $p = .58$; normed fix index (NFI) = .98, comparative fit index (CFI) = 1.0, root mean square error of

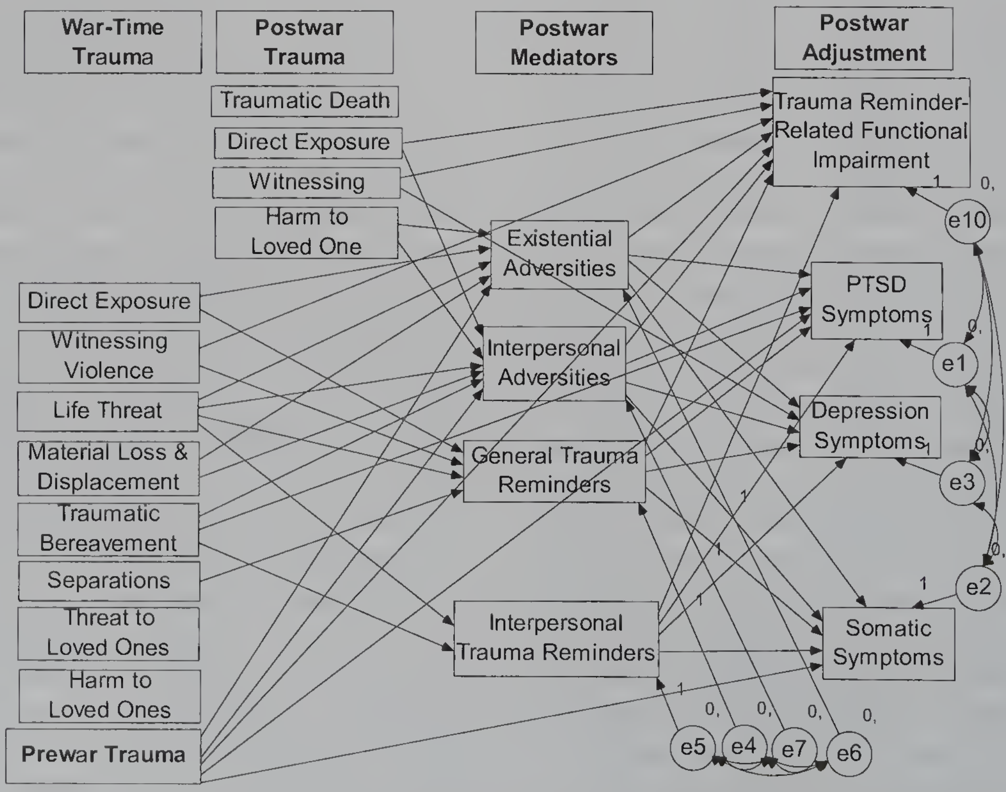

War-Time Trauma
Postwar Trauma
Postwar Mediators
Postwar Adjustment
Traumatic Death
Direct Exposure
Witnessing
Harm to Loved One
Existential Adversities
Interpersonal Adversities
General Trauma Reminders
Interpersonal Trauma Reminders
Trauma Reminder-Related Functional Impairment
PTSD Symptoms
Depression Symptoms
Somatic Symptoms
Direct Exposure
Witnessing Violence
Life Threat
Material Loss & Displacement
Traumatic Bereavement
Separations
Threat to Loved Ones
Harm to Loved Ones
Prewar Trauma
e10
e1
e3
e2
e5
e4
e7
e6
1
0,

Figure 2. Observed variable path analysis model of *composite–causal indicator*-derived dimensions of trauma exposure.

approximation (RMSEA) = 0 (90% confidence interval [CI] = 0–0.02), Hoelter (.01) = 676; and explained 55%, 46%, 27%, and 28% of the variance in PTSD symptoms, reminder-related functional impairment, depression, and somatic symptoms, respectively. Contrary to expectation, *Harm to Loved Ones* and *Threat to Loved Ones* did not exert significant predictive effects. Consistent with Hypothesis 3, the partially mediated model fit significantly better than its fully mediated counterpart, $\chi^2(8) = 55.04$, $p < .01$, indicating that prewar and some "traumatogenic" war-time exposures exerted direct predictive effects. Direct paths linked *prewar trauma* to PTSD symptoms, somatic symptoms, and functional impairment; war-time *traumatic bereavement* to PTSD symptoms and impairment; and war-time *witnessing violence* to impairment. Postwar *direct exposure* had a direct path to impairment, and postwar *witnessing violence* had direct paths to both impairment and depression.

The common factor–effect indicator model for Subsample 1 (see Figure 3), also had good fit $\chi^2(60) = 57.71$, $p = .56$; NFI = .98, CFI = 1.0, RMSEA = 0 (90% CI = 0–0.03), Hoelter (.01) = 705; and explained 55%, 46%, 28%, and 29% of the variances in PTSD symptoms, reminder-related functional impairment, depression symptoms, and somatic symptoms, respectively. Contrary to expectation, neither *Detention & Maltreatment* nor *Personal Assault or Injury* (derived from Factors 3 and 5, respectively) exerted significant predictive effects. Consistent with Hypothesis 3, the partially mediated model fit significantly better than the fully mediated model, $\chi^2(6) = 43.79$, $p < .01$. As in the composite model, *prewar* trauma had three direct paths to PTSD symptoms, somatic symptoms, and functional impairment. Notably, only *Harm to or Death of Loved One* (Factor 4) exerted a direct predictive effect (to depression) whereas the predictive effects of all other war-time variables were fully mediated. *Postwar Direct Exposure* had a direct path to reminder-related functional impairment, and *Postwar Harm to Loved Ones* had a direct path to PTSD symptoms.

In a fourth step, given the susceptibility of "repeated pruning" to sample-idiosyncratic over-fitting (Byrne, 2001), the goodness of fit of both models derived using Subsample 1 was tested (without further modification) in Subsample 2. The composite model had generally good fit, $\chi^2(79) = 159.58$, $p < .01$; NFI = .95, CFI = .97, RMSEA = .05 (90% CI = 0.04–0.06), Hoelter (.01) = 292. The composite model retained (at a significance level of $p < .05$) 28 (68%) of the 41 significant paths identified in Subsample 1 and explained 50%, 53%, 33%, and 30% of the variances in PTSD symptoms, reminder-related functional impairment, depression symptoms, and somatic symptoms, respectively. The common factor model also had generally good fit, $\chi^2(60) = 115.93$, $p < .01$; NFI = .96, CFI = .98, RMSEA

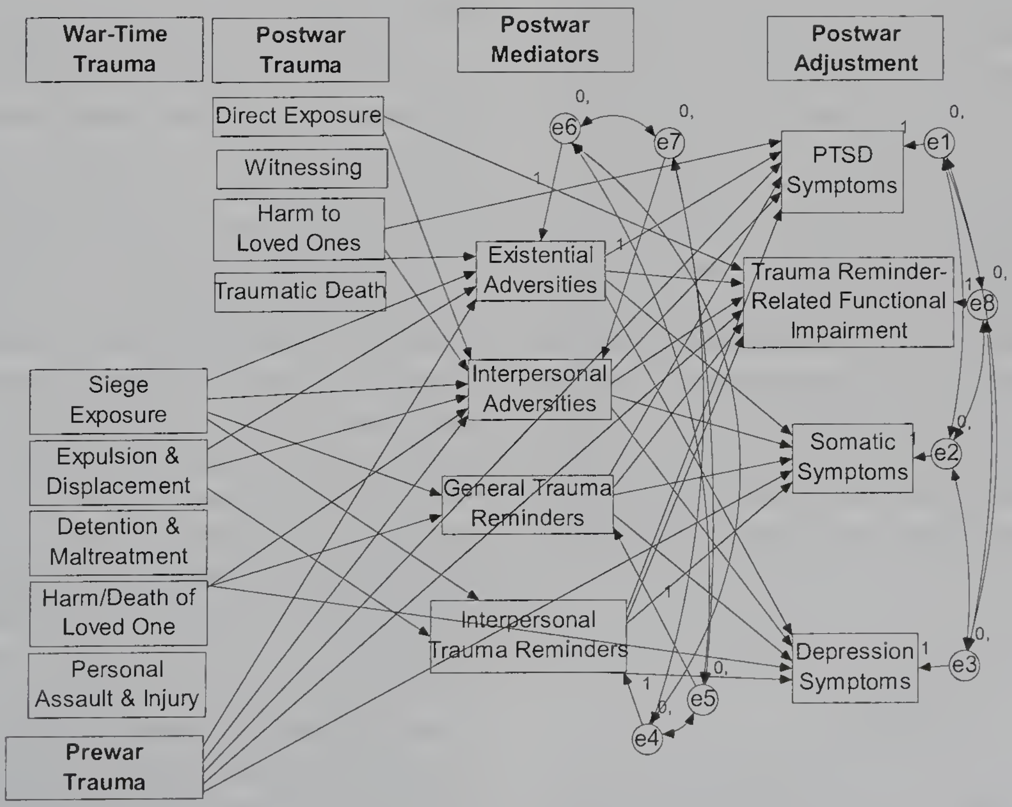

War-Time Trauma
Postwar Trauma
Postwar Mediators
Postwar Adjustment
Direct Exposure
Witnessing
Harm to Loved Ones
Traumatic Death
Siege Exposure
Expulsion & Displacement
Detention & Maltreatment
Harm/Death of Loved One
Personal Assault & Injury
Prewar Trauma
Existential Adversities
Interpersonal Adversities
General Trauma Reminders
Interpersonal Trauma Reminders
PTSD Symptoms
Trauma Reminder-Related Functional Impairment
Somatic Symptoms
Depression Symptoms
e1
e2
e3
e4
e5
e6
e7
e8
0,
1

Figure 3. Observed variable path analysis model of *common factor–effect indicator*-derived dimensions of trauma exposure.

= .05 (90% CI = 0.03–0.06), Hoelter (.01) = 320. The common factor model retained 26 (72%) of the 36 paths in Subsample 1 and explained 51%, 53%, 33%, and 31% of the variances in PTSD symptoms, functional impairment, depression symptoms, and somatic symptoms, respectively. Reduced power (due to halving the sample), as well as restricted predictor variance, likely played roles in reducing the number of significant paths in both Subsample 2 models: In the composite model, 6 of 13 nonsignificant paths had predictor means < 0.10, and 10 of 13 had means < 0.15. In the common factor model, 4 of 10 nonsignificant paths had predictor means < 0.10, and 5 of 10 had means < 0.15.

In a fifth step, we tested Hypothesis 4 (beginning with Subsample 1) using effects decomposition (Byrne, 2001) to rank prewar, war-time, and postwar trauma risk factors by their total predictive effects on each outcome. Total effects for each risk factor were calculated by summing its direct effects (via direct paths) and indirect effects (via mediated paths). The top and bottom halves of Table 3 contain the risk factor rankings for the composite- and common factor-based approaches, respectively. Consistent with Hypothesis 4, the rankings for the composite model showed more differentiation than the common factor model between war-time trauma exposure and outcomes. Four war-time variables (*Life Threat*, *Traumatic Death*, *Witnessing Violence*, and *Loss and Displacement*) appeared in the top six rankings across outcomes in the composite model. In contrast, three war-time variables (*Siege Exposure*, *Expulsion & Displacement*, and *Harm to or Death of Loved Ones*) appeared in the top six rankings for the common factor model; *Siege Exposure* was the dominant risk factor for all outcomes. In a Subsample 2 replication analysis, the risk factor rankings in both models remained relatively stable (rankings not shown). Five war-time variables (the four from Subsample 1 plus *Separation From Loved Ones*) appeared in the top six rankings of the composite model. The common factor model contained the three war-time risk factors found in Subsample 1; *Siege Exposure* was again the dominant risk factor. *Prewar Trauma* was a predictor for all outcomes in both the composite and common factor models, in both Subsamples 1 and 2.

Discussion

This study addressed the question of how to conceptualize and model the multidimensional features of war exposure in ways that best facilitate theory building, strengthen methodological rigor, and improve risk screening and intervention. We presented two different approaches to modeling trauma exposure, the traditionally used *common factor with effect indicators model* and the *composite–causal indicators model*, and discussed their respective strengths in promoting these three aims. We focused on the core assumption of indicator homogeneity (i.e., are trauma exposure indicators *causal effects* of a common factor, or instead *contributors* to a composite?) and its far-reaching implications for how war exposure is conceptualized, measured, studied, and treated. The central question is: *Should specific types of trauma and loss be aggregated into dimensions according to their covariance structure, or instead, according to their consequences?* Given its widespread use via EFA and CFA, we reviewed limitations of the common factor model for applications that aim to model differential links between the occurrence of specific types of traumatic events and specific causal consequences. We posed a caveat: Applications that violate the assumption of indicator homogeneity (shelling cities *causes* sieges, not the reverse) can create "model–application" mismatches. Such mismatches can impede a call to move beyond identifying general "markers of risk" that only denote *who* may need *some form* of mental health services, to instead creating tools that can accurately predict *who* may need *which* types of mental health treatment components across different war and disaster settings (APA Presidential Task Force, 2008).

We further proposed that a *composite*-based approach that "packs" indicators into equifinal war exposure dimensions may better elucidate the differential effects of specific types of trauma on specific outcomes. A composite approach conceptualizes covariance between types of trauma exposure as reflecting *perceptions of event co-occurrence within specific settings*. Shedding unnecessary constraints linked to homogeneity (e.g., retaining only items that load a common factor and that increase "internal consistency") encourages an increased focus on equifinality through aggregating *specific types* of causal risk factors (e.g., traumatic bereavement) according to their common causal effects (PTSD, traumatic grief), and on identifying key *mediators* of those effects (trauma reminders, poverty, family strife).

We compared the common factor–effect indicator model (derived through EFA) to a rationally derived composite–causal indicator model in a split-sample replication design. Results generally supported the four study hypotheses and illus-

Table 3

Rankings (by Predictive Effect) of Trauma Exposure Risk Factors and Postwar Mediators on Posttraumatic Stress, Depression, Somatic Symptoms, and Trauma Reminder-Related Functional Impairment

Composite–Causal Indicator Model of War Trauma Exposure: Risk Factor Rankings by Outcome[a]

Trauma exposure risk factor ranking[b]	Posttraumatic stress (R^2 = 55%) *Risk factor* (total) [direct] {indirect} effect	Depression (R^2 = 27%) *Risk factor* (total) [direct] {indirect} effect	Somatic distress (R^2 = 28%) *Risk factor* (total) [direct] {indirect} effect	Functional impairment (R^2 = 46%) *Risk factor* (total) [direct] {indirect} effect
1	*War-Time Life Threat* (.26) [.00] {.26}	*War-Time Life Threat* (.16) [.00] {.16}	*Prewar Trauma* (.18) [.12] {.06}	*War-Time Life Threat* (.23) [.00] {.23}
2	*Prewar Trauma* (.16) [.12] {.04}	*War-Time Loss & Displacement* (.09) [.00] {.09}	*War-Time Life Threat* (.16) [.00] {.16}	*Prewar Trauma* (.14) [.11] {.03}
3	*War-Time Traumatic Death* (.10) [.08] {.03}[c]	*Prewar Trauma* (.07) [.00] {.07}	*War-Time Loss & Displacement* (.07) [.00] {.07}	*War-Time Witnessing* (.13) [.08] {.05}
4	*War-Time Witnessing* (.07) [.00] {.07}	*Postwar Harm to Loved Ones* (.07) [.00] {.07}	*Postwar Harm to Loved Ones* (.05) [.00] {.05}	*Postwar Direct Trauma Exposure* (.11) [.09] {.02}
5	*War-Time Loss & Displacement* (.06) [.00] {.06}	*Postwar Direct Trauma Exposure* (.04) [.00] {.04}	*War-Time Witnessing Violence* (.03) [.00] {.03}	*War-Time Traumatic Death* (.05) [.00] {.05}
6	*Postwar Harm to Loved Ones* (.05) [.00] {.05}	*War-Time Witnessing* (.02) [.00] {.02}	*Postwar Direct Trauma Exposure* (.02) [.00] {.02}	*War-Time Loss & Displacement* (.04) [.00] {.04}

Postwar mediator ranking[d,e]	*Mediators* [direct effect]	*Mediators* [direct effect]	*Mediators* [direct effect]	*Mediators* [direct effect]
1	*General Trauma Reminders* [.41]	*Interpersonal Adversities* [.24]	*Existential Adversities* [.21]	*General Trauma Reminders* [.31]
2	*Interpersonal Trauma Reminders* [.24]	*Existential Adversities* [.23]	*General Trauma Reminders* [.20]	*Interpersonal Trauma Reminders* [.30]
3	*Interpersonal Adversities* [.17]	*Interpersonal Trauma Reminders* [.17]	*Interpersonal Trauma Reminders* [.15]	*Existential Adversities* [.11]
4	*Existential Adversities* [.14]	*General Trauma Reminders* [.13]	*Interpersonal Adversities* [.13]	*Interpersonal Adversities* [.09]

Common Factor–Effect Indicator Model of War Trauma Exposure: Risk Factor Rankings by Outcome[a]

Trauma exposure risk factor ranking[b]	Posttraumatic stress (R^2 = 55%) *Risk factor* (total) [direct] {indirect} effect	Depression (R^2 = 28%) *Risk factor* (total) [direct] {indirect} effect	Somatic distress (R^2 = 29%) *Risk factor* (total) [direct] {indirect} effect	Functional impairment (R^2 = 46%) *Risk factor* (total) [direct] {indirect} effect
1	*War-Time Siege Exposure* (.28) [.00] {.28}	*War-Time Siege Exposure* (.18) [.00] {.18}	*War-Time Siege Exposure* (.18) [.00] {.18}	*War-Time Siege Exposure* (.24) [.00] {.24}
2	*Prewar Trauma* (.15) [.10] {.04}	*War-Time Expulsion & Displacement* (.13) [.00] {.13}	*Prewar Trauma* (.18) [.12] {.05}	*Prewar Trauma* (.15) [.12] {.03}
3	*War-Time Expulsion & Displacement* (.11) [.00] {.11}	*Prewar Trauma* (.07) [.00] {.07}	*War-Time Expulsion & Displacement* (.11) [.00] {.11}	*Postwar Direct Trauma Exposure* (.10) [.09] {.01}

Table 3
Continued

Common Factor–Effect Indicator Model of War Trauma Exposure: Risk Factor Rankings by Outcome[a]

Trauma exposure risk factor ranking[b]	Posttraumatic stress (R^2 = 55%) *Risk factor* (total) [direct] {indirect} effect	Depression (R^2 = 27%) *Risk factor* (total) [direct] {indirect} effect	Somatic distress (R^2 = 28%) *Risk factor* (total) [direct] {indirect} effect	Functional impairment (R^2 = 46%) *Risk factor* (total) [direct] {indirect} effect
4	*Postwar Harm to Loved Ones* (.10) [.05] {.04}	*Postwar Harm to Loved One* (.07) [.00] {.07}	*Postwar Harm to Loved Ones* (.05) [.00] {.05}	*War-Time Expulsion & Displacement* (.08) [.00] {.08}
5	*Postwar Direct Trauma Exposure* (.03) [.00] {.03}	*Postwar Direct Trauma Exposure* (.04) [.00] {.04}	*Postwar Direct Trauma Exposure* (.02) [.00] {.02}	*Postwar Harm to Loved One* (.03) [.00] {.03}
6	*War-Time Harm to or Death of Loved One* (.01) [.00] {.01}	*War-Time Harm to or Death of Loved One* (−.09) [−.08] {−.01}[f]	*War-Time Harm to or Death of Loved One* (.00) [.00] {.00}	*War-Time Harm to or Death of Loved One* (.02) [.00] {.02}
Postwar mediator ranking[g]	*Mediators* [direct effect]	*Mediators* [direct effect]	*Mediators* [direct effect]	*Mediators* [direct effect]
1	*General Trauma Reminders* [.33]	*Existential Adversities* [.24]	*Existential Adversities* [.21]	*General Trauma Reminders* [.33]
2	*Interpersonal Trauma Reminders* [.30]	*Interpersonal Adversities* [.22]	*General Trauma Reminders* [.20]	*Interpersonal Trauma Reminders* [.30]
3	*Existential Adversities* [.12]	*Interpersonal Trauma Reminders* [.18]	*Interpersonal Trauma Reminders* [.15]	*Existential Adversities* [.12]
4	*Interpersonal Adversities* [.08]	*General Trauma Reminders* [.13]	*Interpersonal Adversities* [.13]	*Interpersonal Adversities* [.08]

Note. R^2 = proportion of variance explained by model.
[a] Predictive effects were calculated using standardized models in which all outcomes and mediators were treated as continuous variables. [b] Six highest-ranking trauma exposure risk factors (by total predictive effect) for each outcome. [c] Direct + indirect effects versus total effects do not always correspond due to rounding error. [d] R^2 values for mediators in composite–causal indicator model = 24% for *General Trauma Reminders*, 12% for *Interpersonal Trauma Reminders*, 27% for *Existential Adversities*, and 19% for *Interpersonal Adversities*. [e] All four of the modeled putative postwar mediators are presented, ranked by direct predictive effect on the outcome, for both models. [f] Negative association suggests a suppressor effect. [g] % R^2 values for mediators in common factor–effect indicator = 19% for *General Trauma Reminders*, 7% for *Interpersonal Trauma Reminders*, 28% for *Existential Adversities*, and 20% for *Interpersonal Adversities*.

trated the relative strengths (and some potential drawbacks) of the two methods. The common factor model extracted clusters of co-occurring events that reflected idiosyncratic features of how the Bosnian conflict was waged (e.g., sieges, ethnic cleansing). The choice of test construction method is thus consequential, as the two methods derived dimensions of war exposure that differed substantially in their elements and relations with theorized mediators and outcomes. Both methods produced good model fit and explained substantial proportions of variance in outcomes. The partially mediated models fit better than their fully mediated counterparts, providing support for the hypothesis that "highly traumatogenic" prewar and war-time risk factors would exert direct predictive effects. However, most predictive effects were transmitted via postwar mediators (trauma reminders and adversities). The composite–causal indicator model produced clearer evidence of differential relations linking prewar and war-time trauma to postwar outcomes. In contrast, the common factor–effect indicator model showed less differentiation, as illustrated by the marked diversity in types of war exposure loading the Siege factor and its dominance as the primary risk factor for all outcomes.

We now discuss the implications of these findings by addressing five questions relating to theory building, strengthening methodological rigor, and improving risk screening, triage, and intervention.

Study Implications: Five Questions Regarding Theory Building, Research Methods, and Intervention

Question 1: Which measurement approach is preferable for modeling war exposure? The good fit of the Subsample 2 composite model is consistent with the assertion that specific types of trauma and loss (e.g., traumatic deaths) may generate similar consequences (e.g., PTSD) without necessarily sharing the same causal origin (i.e., processes underlying a common factor that *cause* the indicators to occur and co-occur) or manifesting "evidence" of indicator homogeneity. Our results suggest that which method is preferable depends on which part of the model—the *structure of co-occurrences among traumatic events*, versus the *structure of paths that differentially link the occurrences of specific types of trauma and loss to specific causal consequences*—is of primary interest. The "big picture" focus of the common factor model makes it well suited for "first-pass" questions that arise in the *situation analysis* phase of crisis response: "How was the war waged—which clusters of events co-occurred?" and "Who is at (global) risk due to the accumulation of co-occurring risk factors?" In contrast, the composite–causal indicator model is more useful for addressing deeper questions concerning *which types of trauma and loss place which subgroups at risk for which outcomes via which pathways of influence*. Such issues are more relevant to the needs assessment, risk screening/triage, case conceptualization, intervention planning, program evaluation, and prevention phases of response (Layne, Beck, et al., 2009) given that they address questions such as: "How should we prioritize trauma types according to severity?" "How can we use limited mental health resources most wisely?" and "Which mediating pathways should our interventions target and seek to interrupt?"

Our results also raise a caveat: "Packing" trauma exposure risk factors into dimensions based on an incorrect assumption regarding their causal origin (i.e., indicator homogeneity) may undermine efforts to "unpack" and clarify their differential relations with their respective causal consequences. Understanding that inter-item covariances *reflect event co-occurrences, not item homogeneity*, is the key to interpreting our puzzling results: The inverse loadings of *mother killed* and *sibling killed* on *Expulsion & Displacement* (Factor 2) reflect the grim strategy of ethnic cleansing—to frighten "undesirable" ethnic groups into fleeing a region by killing those who remain behind. The inverse loading of *left country because of the war* on *Siege* (Factor 1) reflects the difficulties inherent in escaping a siege. The failure of *father killed* to load any factor reflects the widespread conscription of fathers, who were exposed to lethal risks at the front lines that were largely independent of the risks to which their wives and children were exposed. Thus, a serious model–application mismatch may arise when the "real-world" structure among event types is incompatible with the assumptions of indicator homogeneity and of indices often used to gauge homogeneity (e.g., alpha; see Sijtsma, 2009). Although dropping inversely loaded items may drive up estimates of "internal consistency," it reduces test validity by reducing the correspondence between the test and the structure among variables it purports to measure (Borsboom et al., 2003, 2004). These results illustrate how complex patterns of "real-life" nonhomogeneous events can generate complex factor structures that faithfully mirror them (e.g., *Expulsion & Displacement* reflects both being driven from one's home and having one's mother and siblings *not* killed).

Question 2: Are some types of trauma exposure more potent causal risk factors than others, and if so, can they be prioritized? Table 3 illustrates the potential utility of effects decomposition via SEM for identifying who is at risk, for what outcomes, via which pathways. It is striking that all high-ranking risk factors consisted of events (life threat, traumatic death) consistently identified as traumatogenic (La Greca et al., 2002). These findings suggest that different types of trauma and loss "specialize," to a meaningful extent, in the types of risks they convey and the pathways through which they transmit their effects, permitting them to be ranked by total effect in reference to focal outcomes. This differentiation-focused method may help the field move beyond identifying "at-risk" individuals using ambiguous general markers of risk (APA Presidential Task Force, 2008) to instead stratifying them according to their risks for specific outcomes given their exposures to specific risk factors.

Question 3: How can we build our knowledge base in ways that help us develop risk assessment tools that can accurately predict who will need what types of help, across diverse settings and traumatic events? The composite-based method may hold three important advantages over the common factor approach for theory building and intervention applications. First, it promotes the construction of content-valid assessment instruments *that are tailored to specific settings* because its aim is to capture a census rather than a "best internally consistent sample" of trauma exposure. Second, it unpacks settings into *common denominator constructs* (e.g., life threat, traumatic death) that are more likely to be consistent in their composition and causal effects across diverse settings than event-specific variables (e.g., *Siege*). "Packaging" concepts at the more abstract level of *dimensions of trauma types* enhances the transportability of variables, measures, and findings across diverse settings (APA, 2008). Thus, one does not typically import (at least unmodified) measures across settings but rather, a general conceptual framework of exposure dimensions to serve as a test construction template. "Event tailoring" an exposure measure is guided by a two-step algorithm: *(a) Did exposure to a dimension (e.g., life threat) occur? (b) If yes, how was it manifest in this setting? Generate or modify as many items as needed* (e.g., using qualitative rapid needs assessment; see Betancourt, Speelman, Onyango, & Bolton, 2009) *to maximize content, construct, and ecological validity*. Third, "packaging" findings across studies into a common currency of exposure dimensions will help the field learn more efficiently across diverse events by facilitating cross-pollinating discourse between subliteratures (e.g., war, terrorism, hurricanes) that have been unnecessarily compartmentalized by setting-specific, event-bound measures and variables (La Greca et al., 2002). This will increase the growth rate, generalizability, and practical utility of the knowledge base while permitting meta-analytic studies that treat event type or setting as a moderator (e.g., life threat during torture or rape may be more traumatogenic, or may produce more distressing secondary adversities or trauma reminders, than life threat during a hurricane; Norris et al., 2006).

Question 4: How can these methods promote the study of development in high-risk settings? Findings that prewar trauma exerted direct predictive effects 5 years postwar underscores the potency of adverse childhood events in even highly stressful contexts, and suggests that *when* traumatic events occur may be as consequential as *what* occurs. Moreover, the complementary strengths of the common factor and composite-based models suggest that both models, used in tandem, can "unpack" complex events and their aftermath (Coltman, Devinney, Midgley, & Venaik, 2008). Examples include *unpacking risk factor caravans* and *resource caravans*. These caravans respectively consist of clusters of co-occurring risk and promotive/protective factors (e.g., social support) that "travel" with their host across development (Layne, Beck, et al., 2009). The common factor and related methods are well suited to unpack configurations of risk factors and resources that co-occur within and across developmental stages. Conversely, the composite approach is better suited to unpack the differential relations between individual caravan elements and their respective consequences. Such a dual-pronged approach can support not only secondary and tertiary interventions that seek to reduce psychopathology and dysfunction, but also strength-based public health approaches that emphasize prevention and positive development through the orchestration of resource gain cycles. This dual approach can also help to disentangle both the complex array of risk factors theorized to *cause* developmental trauma disorder—a constellation of cognitive, affective, somatic, and behavioral problems now under consideration for *DSM–V* (Courtois & Gold, 2009)—as well as its structure. The composite model broadly applies (Dodge et al., 2008): "Latent" sieges are no more a *cause* of shelling youth than physical abuse causes beatings, emotional abuse causes screaming, and community violence causes stabbings.

The common factor and composite methods can also be used to study differential risks for exposure to types of trauma as a function of developmental stage. Because adolescents often gather supplies in war settings and have social networks populated with peers and adult friends (Layne, Warren, et al., 2009), they may be at higher risk than children for many types of exposure, including life threat, witnessing injury or death, and having a close friend killed or injured. In contrast, children and adolescents may be at equivalent risk for loss of a parent. Differences in child versus adolescent reactions to war exposure may thus reflect differential exposure *to* specific types of trauma, differential effects *of* those exposures, and maturational and other contextual effects (Pat-Horenczyk et al., 2009; Pynoos et al., 1995; Shaw, 2003).

Question 5: How can these findings improve assessment and intervention? The limited utility of the common factor model for clinical applications is illustrated by the question, "What does knowing that siege exposure is the primary risk factor for every outcome really tell us about how to help war-exposed youth?" *Co-occurring* risk factors (e.g., all indicators loading *Siege*) are risk markers for one another, and for general "at risk" status (APA Presidential Task Force, 2008); thus, they help to identify *who may need help*. However, such risk markers provide very little guidance regarding *how, where, when, and why to help*. To the extent that it disperses equifinal risk factors (e.g., types of traumatic death) across different test dimensions, aggregates causally diverse risk factors into the same dimension (e.g., *Siege*), or drops "uncooperative" critical items (e.g., *father killed*), the common factor method can undermine assessment in four ways. First, causally complex exposure dimensions are ambiguous: Clinicians do not diagnose or treat clients for siege exposure per se. Such "risk factors" are too complex and vague in their makeup to serve any purpose other than general "markers of risk" (APA Presidential Task Force, 2008). Second, ambiguous test scores increase the risk for false negatives and false positives (Kraemer et al., 1999): A youth who is devastated by witnessing his father's death may score 3 on a war exposure inventory (and be overlooked), whereas a youth exposed to many less severe events may score 10. Third, because its "data input" is the structure of covariances *among risk factors*, the common factor model can create "context-specific" variables (e.g., *Siege*) and tests that reflect idiosyncratic ways in which risk factors co-occurred within a given setting and that have limited transportability across settings (APA Presidential Task Force, 2008).

A fourth problem is that causally diverse variables dilute, blur, and obscure, rather than clarify, differential relations between causes and consequences. *Siege Exposure* is the primary risk factor for all outcomes in this study *precisely because it is a complex composite of risk factors (or risk markers) for all outcomes.* Our capacity to accurately predict who will need services (APA Presidential Task Force, 2008) hinges on our ability to differentiate between the potencies, pathways of influence, and consequences of specific causal risk factors. Failure to discriminate adversely affects all "downstream" activities that require knowledge of *who is at risk for what outcomes via which pathways of influence* to make good evidence-based clinical decisions. Models in which most risk factors relate indiscriminately to most outcomes make elimination of *nonsignificant* pathways difficult. This reduction in model parsimony can both increase false-negative rates (by not ruling out those *not* at high risk for specific outcomes) and reduce efficiency (as specialized treatment components are assigned unnecessarily). For example, our results suggest that siege-exposed youths are at highest risk for all outcomes and thus are of higher priority than all other groups and should receive broad-spectrum treatments targeting all outcomes.

In contrast, better differentiated composite–causal indicator models may promote cost-effective, public health-based interventions (Layne, Beck, et al., 2009). Because subgroups most likely need services for those outcomes for which their types of exposure place them at risk, differentiated models can balance the *level of need* (groups at higher risk receive more specialized services), *effectiveness* (necessary services are provided), and *efficiency* (no unnecessary specialized services are provided). Differentiated models help to identify the specific types of trauma and loss that exert the greatest causal influence on each outcome and their respective pathways of influence. This knowledge will shed much-needed light on the key determinants of the clinical course of postwar distress and help to identify, prioritize, and triage youths and families to components that target their difficulties. A major advantage of differentiated models is their capacity to identify *"high-value" targets for intervention within the broader social and physical ecology in ways that facilitate assessment-driven, theoretically grounded, and evidence-based treatment tailoring at the individual case level.* Such an approach will encourage wellness, prevention, and intervention efforts that systematically address youth's specific life circumstances (e.g., orchestrating personal resource gain cycles to build "stress resistance-enhancing" resource caravans) rather than assigning treatment components to individuals solely on the basis of clinical cutoff scores on "end point" distress and dysfunction measures (Layne, Beck, et al., 2009). Subgroups exposed only to "indirect effect" (fully mediated) types of trauma may receive substantial and perhaps sufficient benefit from supportive skills-based components that interrupt key mediators and mitigate key moderators (Layne et al., 2007). Evidence that existential adversities (financial hardships) and trauma reminders (destroyed buildings) mediate the links between war exposure and postwar distress blurs the traditional distinction between "material" versus "psychosocial" postwar interventions: Rebuilding infrastructure to improve access to basic necessities, employment, housing, and schools is also a potent psychosocial intervention—a major added benefit for the same investment (Miller & Rasmussen, 2010). Further, evidence that highly traumatogenic events exert both indirect *and* direct effects suggests ways to introduce supplemental specialized components for high-risk groups for which mediator-focused interventions may be insufficient. For example, Trauma and Grief Component Therapy (Layne et al., 2008) is a multitiered school-based intervention that includes "Tier 1" classroom components targeting key mediators (e.g., coping with trauma reminders) and moderators (e.g., recruiting social support), and "Tier 2" group-based components targeting highly traumatogenic events (trauma and grief processing). A program evaluation (Cox et al., 2007; Layne et al., 2008) found that Tier 1 components were widely disseminated at low cost, producing substantial benefit with low iatrogenic risk. The Tier 2 condition produced improvement rates comparable to rigorously controlled treatment efficacy trials.

Study Limitations

This study used a self-report screening index of life events (Monroe, 2008) and did not include measures of loss reminders (a theorized mediator; Layne et al., 2006) and grief reactions; both measures are included in a forthcoming study. The cross-sectional retrospective design precludes both causal inference (as event reporting may be confounded with current psychological functioning) and the longitudinal analysis of adjustment trajectories (Layne, Beck, et al., 2009). The extended time frame (5 years postwar) created a window in which other risk factors (especially postwar trauma) could occur and compete for explained variance, allowing perhaps only the most traumatogenic or impactful war-time risk factors to exert significant predictive effects. The time frame also sampled youths 8–14 years old at the *end* of the war—a group less likely to be severely exposed than older adolescents. This may have led to low endorsement rates for some event types (e.g., physical injury) that prevented their predictive effects from being adequately tested. In addition, the purposive sampling of classes with a high prevalence of severely war-exposed students limits the generalizability of findings to more severely exposed youths residing in the region. Moreover, the variance decomposition method used to rank risk factors assumes that all nontrivial variables are included in the model; omitting influential variables may alter parameter estimates and risk factor rankings (Byrne, 2001). Further study is also needed to evaluate the divergence in performance between the two approaches, as well as the generalizability and utility of the eight-dimensional framework used in the composite approach, in other settings and applications. Last, although evidence for its utility is accumulating (Coltman et al., 2008), critics (e.g., Wilcox et al., 2008) question the applicability of the composite model; for example, different research teams could produce different sets of indicators purporting to measure the same composite. We concur that this is possible but assert that the caveat applies with equal force to the common factor model, which may produce "researcher-idiosyncratic" discrepancies arising from item wordings, guiding theory, and so forth.

References

Ahmetasevic, N. (2007, June 20). *Bosnia's book of the dead.* Justice Report, Balkan Investigative Reporting Network. Retrieved May 10, 2010, from http://birn.eu.com/en/88/10/3377/.

APA Presidential Task Force on Posttraumatic Stress Disorder and Trauma in Children and Adolescents. (2008). *Children and trauma: Update for mental health professionals.* Washington, DC: American Psychological Association.

Barenbaum, J., Ruchkin, V., & Schwab-Stone, M. (2004). The psychosocial aspects of children exposed to war: Practice and policy initiatives. *Journal of Child Psychology and Psychiatry, 45,* 41–62.

Betancourt, T. S., Speelman, L., Onyango, G., & Bolton, P. (2009). A qualitative study of mental health problems among children displaced by war in Northern Uganda. *Transcultural Psychiatry, 46,* 238–256.

Birelson, P. (1981). The validity of depressive disorder in childhood and development of a self-rating scale: A research project. *Journal of Clinical Psychology and Psychiatry, 22,* 73–88.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110,* 305–314.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110,* 203–219.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061–1071.

Byrne, B. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* New York: Psychology Press.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41,* 189–224.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61,* 1250–1262.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). *Hillsdale,* NJ: Erlbaum.

Courtois, C. A., & Gold, S. N. (2009). The need for inclusion of psychological trauma in the professional curriculum: A call to action. *Psychological Trauma: Theory, Research, Practice, and Policy, 1,* 3–23.

Cox, J., Davies, D. R., Burlingame, G. M., Campbell, J. E., Layne, C. M., & Katzenbach, R. J. (2007). Effectiveness of a trauma/grief-focused group intervention: A qualitative study with war-exposed Bosnian adolescents. *International Journal of Group Psychotherapy, 57,* 319–345.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61,* 1203–1218.

Dodge, K. A., Greenberg, M. T., Malone, P. S., & Conduct Problems Prevention Research Group (2008). Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development, 79*, 1907–1927.

Dyregrov, A., Gupta, L., Gjestad, R., & Mukanoheli, E. (2000). Trauma exposure and psychological reactions to genocide among Rwandan children. *Journal of Traumatic Stress, 13*, 3–21.

Entholt, K., & Yule, W. (2006). Practitioner review: Assessment and treatment of refugee children and adolescents who have experienced war-related trauma. *Journal of Child Psychology and Psychiatry, 47*, 1197–1210.

Fayyad, J., Karam, E., Karam, A. N., Tabet, C. C., & Ghosn, M. B. (2004). PTSD in children and adolescents following war. In R. R. Silva (Ed.), *Posttraumatic stress disorders in children and adolescents* (pp. 306–352). New York: W.W. Norton.

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research, 61*, 1229–1237.

Goodman, L. A., Corcoran, C., Turner, K., Yuan, N., & Green, B. L. (1998). Assessing traumatic event exposure: General issues and preliminary findings for the Stressful Life Events Screening Questionnaire. *Journal of Traumatic Stress, 11*, 521–542.

Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics, 15*, 191–213.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1999). Measuring the potency of risk factors for clinical or policy significance. *Psychological Methods, 4*, 257–271.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56*, 921–926.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (2002). *Helping children cope with disasters and terrorism*. Washington, DC: APA.

Layne, C. M. (1999a). *Pre-War Trauma Exposure Index*. Unpublished psychological test, University of California, Los Angeles.

Layne, C. M. (1999b). *Post-War Trauma Exposure Index*. Unpublished psychological test, University of California, Los Angeles.

Layne, C. M., Beck, C. J., Rimmasch, H., Southwick, J. S., Moreno, M. A., & Hobfoll, S. E. (2009). Promoting "resilient" posttraumatic adjustment in childhood and beyond: "Unpacking" life events, adjustment trajectories, resources, and interventions. In D. Brom, R. Pat-Horenczyk, & J. Ford (Eds.). *Treating traumatized children: Risk, resilience, and recovery* (pp. 13–47). New York: Routledge.

Layne, C. M., & Djapo, N. (1999). *Post-War Adversities Index*. Unpublished psychological test, University of California, Los Angeles.

Layne, C. M., Saltzman, W. R., Poppleton, L., Burlingame, G. M., Pasalic, A., Durakovic, E., et al. (2008). Effectiveness of a school-based group psychotherapy program for war-exposed adolescents: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 1048–1062.

Layne, C. M., Steinberg, A. M., & Pynoos, R. S. (1998). *Trauma Reminders Screening Inventory*. Unpublished psychological test, University of California, Los Angeles.

Layne, C. M., Stuvland, R., Saltzman, W. R., Djapo, N., & Pynoos, R. S. (1999). *War Trauma Exposure Index*. Unpublished psychological test, University of California, Los Angeles.

Layne, C. M., Warren, J. S., Hilton, S., Lin, D., Pasalic, A., Fulton, J., et al. (2009). Measuring adolescent perceived support amidst war and disaster: The Multi-Sector Social Support Inventory. In B. K. Baker (Ed.), *Adolescents and war: How youth deal with political violence* (pp. 145–176). New York: Oxford University Press.

Layne, C. M., Warren, J. S., Saltzman, W. R., Fulton, J. B., Steinberg, A. M., & Pynoos, R. S. (2006). Contextual influences on posttraumatic adjustment: Retraumatization and the roles of revictimization, posttraumatic adversities, and distressing reminders. In L. A. Schein, H. I. Spitz, G. M. Burlingame, P. R. Muskin, & S. Vargo (Eds.), *Psychological effects of catastrophic disasters: Group approaches to treatment* (pp. 235–286). New York: Haworth.

Layne, C. M., Warren, J. S., Watson, P. J., & Shalev, A. Y. (2007). Risk, vulnerability, resistance, and resilience: Toward an integrative conceptualization of posttraumatic adaptation. In M. J. Friedman, T. M. Keane, & P. A. Resnick (Eds.), *Handbook of PTSD: Science and practice* (pp. 497–520). New York: Guilford.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192–211.

MacDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*, 710–730.

Macksoud, M. S., & Aber, J. L. (1996). The war experiences and psychosocial development of children in Lebanon. *Child Development, 67*, 70–88.

Masten, A. S., Neemann, J., & Andenas, S. (1994). Life events and adjustment in adolescents: The significance of event independence, desirability, and chronicity. *Journal of Research on Adolescence, 4*, 71–97.

Miller, K. E., & Rasmussen, A. (2010). War exposure, daily stressors, and mental health in conflict and post-conflict settings: Bridging the divide between trauma-focused and psychosocial frameworks. *Social Science & Medicine, 70*, 7–16.

Monroe, S. M. (2008). Modern approaches to conceptualizing and measuring human life stress. *Annual Review of Clinical Psychology, 4*, 33–52.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. New York: Taylor & Francis.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 81–117.

Muthén, B. O., & Muthén, L. (2009). *Mplus 5.1*. Los Angeles, CA: Author.

Netland, M. (2001). Assessment of exposure to political violence and other potentially traumatizing events. *Journal of Traumatic Stress, 14*, 311–326.

Netland, M. (2005). Event-list construction and treatment of exposure data in research on political violence. *Journal of Traumatic Stress, 18*, 507–517.

Norris, F. H., Galea, S., Friedman, M. J., & Watson, P. J. (Eds.). (2006). *Methods for disaster mental health research*. New York: Guilford.

Pat-Horenczyk, R., Qasrawi, R., Lesack, R., Haj-Yahia, M., Peled, O., Shaheen, M., et al. (2009). Posttraumatic symptoms, functional impairment, and coping among adolescents on both sides of the Israeli–Palestinian conflict: A cross-cultural approach. *Applied Psychology: An International Review, 58*, 688–708.

Powell, S., & Duraković-Belko, E. (Eds.). (2002). *Sarajevo 2000: The psychosocial consequences of war. Results of empirical research from the territory of former Yugoslavia*. Retrieved May 10, 2010, from http://psih.org/2000e.pdf.

Pynoos, R. S., Steinberg, A. M., & Wraith, R. (1995). A developmental model of childhood traumatic stress. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation. Wiley Series on personality processes* (pp. 72–95). Oxford, England: Wiley.

Richters, J. E. (1997). The Hubble hypothesis and the developmentalist's dilemma. *Development and Psychopathology, 9*, 193–229.

Shaw, J. A. (2003). Children exposed to war/terrorism. *Clinical Child and Family Psychology Review, 6*, 237–246.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107–120.

SPSS. (2009). *Amos 16.0 (Software)*. Chicago: Author.

Steinberg, A. M., Brymer, M., Decker, K., & Pynoos, R. S. (2004). The UCLA PTSD Reaction Index. *Current Psychiatry Reports, 6*, 96–100.

Wells, M., Burlingame, G., Lambert, M., Hoag, M., & Hope, C. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and the Youth Outcome Questionnaire. *Psychotherapy, 33*, 275–283.

Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research, 61*, 1219–1228.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1077–1095

Sierra Leone's Former Child Soldiers: A Follow-Up Study of Psychosocial Adjustment and Community Reintegration

Theresa Stichick Betancourt,
Ivelina Ivanova Borisova, and
Timothy Philip Williams
François-Xavier Bagnoud Center for Health and Human Rights, Harvard School of Public Health

Robert T. Brennan
Harvard Medical School

Theodore H. Whitfield
Independent Biostatistics Consultant

Marie de la Soudiere
Independent Consultant, UNICEF

John Williamson
Displaced Children and Orphans Fund of USAID

Stephen E. Gilman
Harvard School of Public Health

This is the first prospective study to investigate psychosocial adjustment in male and female former child soldiers (ages 10–18; $n = 156$, 12% female). The study began in Sierra Leone in 2002 and was designed to examine both risk and protective factors in psychosocial adjustment. Over the 2-year period of follow-up, youth who had wounded or killed others during the war demonstrated increases in hostility. Youth who survived rape not only had higher levels of anxiety and hostility but also demonstrated greater confidence and prosocial attitudes at follow-up. Of the potential protective resources examined, improved community acceptance was associated with reduced depression at follow-up and improved confidence and prosocial attitudes regardless of levels of violence exposure. Retention in school was also associated with greater prosocial attitudes.

At present, Sierra Leone is still contending with the legacy of the 1991–2002 civil conflict involving the Revolutionary United Front (RUF), the Sierra Leonean Armed Forces Revolutionary Council (AFRC), the Sierra Leone Army, and local groups like the Civil Defense Forces (CDF). This bloody civil war led to human rights abuses, including mass mutilations and the pervasive use of children in armed conflict. Thousands of children, some as young as 7 years old, were conscripted into fighting

---

We are endlessly grateful to all the local research assistants who carried out these interviews in Sierra Leone and our field-based project coordinators, Moses Zombo and Nassrin Farzaneh, who supervised the teams in 2002 and 2004. We are extremely indebted to Lloyd Feinburg, John Williamson, and Cathy Savino of USAID's Displaced Children and Orphans Fund (DCOF). In addition, we are grateful to our colleagues at the International Rescue Committee, including Jane Warburton, Jodi Nelson, and many others who helped this project to succeed along the way. In particular, we would like to thank Catherine Weisner whose leadership was essential for launching the first wave of data collection. We would like to thank Laura Nolan, Kate Ittleman, and Harlyn Sidhu who helped compile and prepare the tables presented in this article and Julia Rubin-Smith who coordinated our Boston-based team. We would also like to thank Dr. Maggie Alegria of Harvard Medical School for her review of previous versions of this article, Dr. Garrett Fitzmaurice of McLean Hospital for his excellent statistical advice, and Sidney Atwood and Kathy McGaffigan for their tremendous skill in data management and analyses. This work was funded by USAID's Displaced Children and Orphan's Fund. The present publication was also supported by Grant 1K01MH077246-01A2 from the National Institute of Mental Health and Grant P60 MD002261 from the National Center for Minority Health and Disparities.

Correspondence concerning this article should be addressed to Theresa Stichick Betancourt, Department of Global Health and Population/François-Xavier Bagnoud Center for Health and Human Rights, Harvard School of Public Health, 651 Huntington Avenue, 7th Floor, Boston, MA 02115. Electronic mail may be sent to theresa_betancourt@harvard.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0005

forces and paramilitary groups (Betancourt, Simmons, et al., 2008; The World Revolution, 2001). Many young boys and girls were brutalized into submission, drugged, and then forced to fight on the frontlines of combat; others who did not fight served as porters, cooks, guards, messengers, servants, human shields, or bush "wives." Although most point to abduction as their point of entry into their role as child soldiers (Mazurana & Carlson, 2004), some children may have assumed a more active role in their decision to participate in armed conflict. The breakdown of family and community systems, coupled with insufficient educational opportunities, left many children with limited viable options for the future (Ashby, 2002). Consequently, a militia group's promise of adventure, security, and camaraderie led some to turn to conscription into armed conflict as their "best among the worst" of options (Machel, 2001; Peters & Richards, 1998; Shepler, 2005). As a result of their involvement with armed forces and armed groups, many youth were witnesses and/or perpetrators of intense physical violence, including summary executions and death-squad killings, torture, detention, rape, bombings, forced displacement, destruction of homes, and massacres of family members. They were deprived of their rights to the care and protection of their families, and denied education and other developmental opportunities.

The Sierra Leone conflict involved direct attempts by rebel groups to destroy relations between young abductees and their families and communities. This was done by forcing many young people to commit atrocities (killings, assault) against loved ones and neighbors under the threat of death. This violent indoctrination into fighting forces was intended to sever ties that might lead abducted youth to attempt escape and return home (Betancourt, Pochan, & de la Soudiere, 2005; Wessells, 2005). Nevertheless, with the signing of the Abuja agreements in November 2000 and May 2001, most of Sierra Leone's former child soldiers began a long journey of return to their families and communities.

Disarmament, demobilization, and reintegration (DDR) refers to the process by which soldiers are returned to civilian life as conflicts end. DDR programs aimed specifically at child soldiers have been established in several countries, including Sierra Leone, where the National Center for DDR implemented a program of demobilization in conjunction with the United Nations Mission in Sierra Leone (UNAMSIL) and the warring parties (RUF, CDF, AFRC). These activities resulted in the demobilization of more than 72,000 combatants—including over 6,700 boys and girls—across the country (National Committee for Demobilisation Disarmament and Reintegration, 2004; United Nations International Children's Emergency Fund, 2005).

The first step in DDR, disarmament, involves soldiers turning in their weapons, which are usually destroyed. The second step, demobilization, involves disbanding armed groups and releasing former child soldiers into civilian society. This step often includes family tracing in order to identify remaining family members or potential alternate caregivers for returning youth. The third phase, reintegration, is a much longer and more complex process. It is in this phase that psychosocial issues take center stage. Given direct attempts by armed forces to destroy relationships between abducted children and their families and communities, social reintegration remains a challenging issue in post-conflict Sierra Leone. To facilitate this process, sensitization campaigns were implemented to initiate local discussions about the reintegration of former child soldiers in the community. In addition, upon reunification, efforts were made to place children back into intact families, particularly families involving biological relatives (Betancourt, Pochan, et al., 2005; Williamson, 2005). Although many organizations identified and trained community volunteers to provide monitoring and individual follow-up of resettled former child soldiers, no formal research has examined the social reintegration and psychosocial adjustment of these youth over time. This study is intended to address these gaps.

Child Soldiers: A Global View

Despite a near universal condemnation from the international community (United Nations International Children's Emergency Fund, 2003), the use of child soldiers remains an ongoing reality in many modern armed conflicts. The 2008 Child Soldiers Global Report documents the abduction and forceful conscription of children (both boys and girls) into government forces and armed groups in at least 86 countries to date (Coalition to Stop the Use of Child Soldiers, 2008). Since 2001, conflicts in Africa have involved the conscription of children into armed forces and armed groups including wars in Angola, Burundi, the Central African Republic, Chad, Cote d'Ivoire, the Democratic Republic of Congo, Guinea, Liberia, Rwanda, Sierra Leone, Somalia, and northern Uganda (Coalition to Stop the Use of Child Soldiers, 2004, 2008).

Child soldiers include boys and girls, many of whom are not fighters but are used in other ways.

For these reasons, the 2007 Paris Principles established the definition of a child associated with an armed force or armed group as "any person below 18 years of age who is or who has been recruited or used by an armed force or armed group in any capacity, including but not limited to children, boys and girls, used as fighters, cooks, porters, messengers, spies or for sexual purposes. It does not only refer to a child who is taking or has taken a direct part in hostilities" (United Nations International Children's Emergency Fund, 2007, p. 7). The Paris Principles and other policy efforts have illuminated the child protection and rehabilitation issues central to addressing the problem of child soldiers globally. Experts contend that the term *children associated with armed forces and armed groups* (CAAFAG) is more accurate and captures the realities faced by children involved with armed groups and armed forces far beyond the proportion who actually engage as combatants (Betancourt, Borisova, et al., 2008). For the purposes of the present study, we use the term *child soldier* for ease of readability with the caveat that we are implying the broader view involved in the use of the term *CAAFAG*.

Psychosocial Adjustment and Social Reintegration of Former Child Soldiers

Although the research on this topic is scant, war-related exposures and unsuccessful transitions into civilian life are thought to place former child soldiers at increased risk of mental health problems, re-recruitment, and many other forms of exploitation (Betancourt, Borisova, et al., 2008; Betancourt, Pochan, et al., 2005; Eyber & Ager, 2003). Using various measures of psychosocial well-being and mental health, a handful of cross-sectional studies with former child soldiers have reported high levels of psychosocial problems relatively soon after reintegration. For instance, Allen and Schomerus (2006) collected data on a large ($n = 886$) sample of former child soldiers under 18 years in Uganda who had recently reintegrated into their communities. Although the study did not use standardized measures of mental health outcomes, high levels of psychological trauma and social difficulties were reported. Using the Impact of Event Scale–Revised (Weiss & Marmar, 1997), Derluyn, Broekaert, Schuyten, and De Temmerman (2004) documented high levels of symptoms (97%) of posttraumatic stress disorder (PTSD) among 71 former child soldiers in northern Uganda. Among Ugandan and Congolese former child soldiers ($n = 169$), Bayer, Klasen, and Adam (2007) found that more than one third (34.9%) met diagnostic criteria for PTSD. Similarly, Kohrt's (2007) study in Nepal found that children associated with fighting forces exhibited moderately higher psychosocial problems (aggression, depression, anxiety, and posttraumatic stress) when compared with never-associated children.

However, not all studies of former child soldiers document high rates of mental distress or disorders. For instance, Annan, Blattman, and Horton's (2006) Survey of War-Affected Youth (SWAY) in northern Uganda indicated that despite high rates of exposure to violence among male former child soldiers, levels of emotional distress were lower than might otherwise be expected. When compared with children never abducted by fighting forces, psychological problems among returning former child soldiers appeared mild to moderate, with serious distress concentrated in the minority of participants who experienced extreme violence (including perpetration of violence). In addition, the correlation between abduction and aggression was weak or nonexistent (Blattman & Annan, 2007). Using retrospective reports, the SWAY study team found that even though some formerly abducted females reported difficulties with family and community acceptance upon return, a majority reported increased acceptance over time. The authors concluded that abduction itself was generally not a good predictor of mental health problems (Annan, Blattman, Carlson, & Mazurana, 2008).

The studies of former child soldiers reviewed earlier involved data collection on samples relatively soon after demobilization or immediately following community reintegration. Only a handful of studies have assessed mental health and psychosocial functioning over time, addressing the potential long-term psychological consequences of involvement in armed forces and armed groups. Santacruz and Arana (2002) interviewed 239 former child soldiers, who had belonged to armed groups in El Salvador, nearly 10 years after the end of its civil war. At the time of these interviews, a majority of the sample continued to have vivid memories of war-related violence. Twenty percent suffered from constant insomnia and 16% reported persistent nightmares. In addition, 39% said that they felt tired and depressed quite often, 39% reported problems with nervousness, and 37% said they were easily annoyed and angered. This study highlights the potentially enduring difficulties faced by adults who were associated with armed groups as children.

To date, only one prospective longitudinal study of former child soldiers is available in the published literature. Boothby and colleagues followed a small cohort of male former child soldiers from Mozambique over a period of 16 years and outlined some of the long-term effects of child involvement with armed groups (Boothby, Crawford, & Halperin, 2006). Overall, the results of this research suggest that a majority of the 39 young men studied made significant progress in returning to civilian life. However, none were able to fully escape their violent pasts. In 1988, while at the Lhanguene Rehabilitation Center in Mozambique, each young man in the sample reported experiencing recurrent thoughts and memories of traumatic events. Overall, symptoms of distress persisted 16 years later, but the proportion of former child soldiers experiencing problems as adults was considerably lower than those reporting these problems as children (Boothby, 2006). At the last point of assessment, 50% of the participants reported emotional and physical reactions or symptoms indicative of traumatic stress reactions when reminded of hurtful or traumatic events. Additionally, the research indicated that longer length of time spent in the captivity of fighting forces seemed to augment risk of poor long-term social and emotional adjustment (Boothby, 2006). Although the small sample size ($n = 39$) of this study does not allow for in-depth statistical analyses, it does provide an important perspective into the long-term psychosocial consequences of child soldiering.

Former Child Soldiers: An Interactive View of Resilience

One of the questions of greatest interest regarding the well-being of former child soldiers is: How do these youth fare over time? Some surrender to the notion that these youth are a ''lost generation,'' essentially too disturbed and violent to function productively. Some, however, think these youth may represent a subset of resilient individuals who drew from personal strengths and external supports to survive a harrowing experience. A resilience perspective (Layne, Warren, Watson, & Shalev, 2007; Masten & Obradovic, 2008; Rutter, 2006) provides a complex and integrated conceptual framework for thinking about posttraumatic adaptation in former child soldiers over time. Resilience is regarded as an interactive concept referring to ''a relative resistance to environmental risk experiences, or the overcoming of stress or adversity'' (Rutter, 2006). The traumatic stress literature provides an integrative perspective on resilience and adaptation that is helpful for understanding the mental health of young people exposed to acute traumatic events such as armed conflict (Layne et al., 2007). Within the overarching framework of resilience, theories of *stress accumulation* and *stress adjustment* offer complementary views for understanding former child soldiers' response to the compounded adversity characterizing their experience.

Theories of *stress accumulation* posit that the risk for poor mental health and developmental outcomes rises markedly with the accumulation of adverse life experiences (Dohrenwend & Dohrenwend, 1974; Lazarus & Folkman, 1984; Rutter, Pickles, Murray, & Eaves, 2001; Thoits, 1991). A stress accumulation model suggests that children are profoundly affected through repeated and prolonged exposure to severe trauma or a series of stressful events that exceed an individual's threshold level of stress. Such exposure to stressful events can lead to a prolonged activation of biologically based coping strategies that, when repeatedly activated, may compromise individuals' capacity to readjust, leaving them increasingly susceptible to physical or psychological problems throughout the life course (Dohrenwend & Dohrenwend, 1974; McEwen, 2000; Worthman & Panter-Brick, 2008). Particularly, when these events occur early in life, the accumulation of such chronic or ''toxic stressors'' (National Scientific Council on the Developing Child, 2005) may limit the degree to which protective processes may be able to offset exposure to repetitive adversity (Beckett et al., 2006; Nelson et al., 2007; Smyke et al., 2007).

Resilience research has documented a number of protective processes that can help children achieve good functioning, despite exposure to a range of difficult life circumstances (Luthar & Goldstein, 2004). The *stress adjustment paradigm* (Lazarus & Folkman, 1984) emphasizes the role of individual, familial, and community resources in understanding the relation between adverse life events, stressors, and subsequent psychological adjustment and adaptation (Masten & Obradovic, 2008). According to this framework, trauma, psychological adjustment, resilience, and the mental health of former child soldiers must be viewed as a dynamic process involving a number of ecological levels offering potential sources of protection (Bronfenbrenner, 1979) rather than solely individual characteristics (Betancourt & Khan, 2008; Lazarus & Folkman, 1984; Masten & Obradovic, 2008).

Resilience theory and, within it, theories of stress accumulation and stress adjustment direct us to consider the complex interplay of risk and protective processes (Rutter, 2006), both during and after conflict, that likely shape the psychosocial outcomes and adjustment trajectories of former child soldiers. Young people associated with fighting forces may differ with regard to the frequency, duration, and severity of their war experiences, as well as the accumulation of war-related stressors they experienced. Former child soldiers may also differ with regard to their postwar situations, particularly in terms of the family and community supports and resources available to them. Therefore, it is important to consider the accumulation and combination of both risk and protective factors in the lives of former child soldiers to better understand their long-term psychosocial adjustment (Betancourt, Borisova, et al., 2008; Wessells, 2006).

On the basis of the existing evidence cited earlier, we are able to develop hypotheses regarding the influences of specific war-related exposures and potential postconflict protective factors on the psychosocial well-being of former child soldiers. Exposure to violence is the most common risk during wartime (Annan et al., 2006; Bayer et al., 2007; Kohrt et al., 2008). Although most children affected by war experience high rates of exposure to violence (Barenbaum, Ruchkin, & Schwab-Stone, 2004; Lustig et al., 2004), child soldiers are distinguished by experiencing particularly toxic forms of violence, often involving repeated personal victimization and in the most extreme cases, being forced to kill others (Coalition to Stop the Use of Child Soldiers, 2008). For example, in northern Uganda, Derluyn et al. (2004) observed very high rates of violent experiences with a majority of the sample reporting exposure to more than six traumatic events. Seventy-seven percent of the sample reported witnessing the killing of others, 35% reported sexual abuse, and 39% of the sample reporting having wounded or killed others.

Gender is an important factor shaping the war experiences of children associated with armed forces and armed groups. Females comprise a significant proportion of children involved with armed groups globally (McKay & Mazurana, 2004). Although rape occurs among both male and female child soldiers (Annan et al., 2008; Betancourt, Borisova, et al., 2008; Johnson et al., 2008; McKay & Mazurana, 2004), sexual victimization presents particular challenges for psychosocial adjustment and social reintegration as female former child soldiers may be viewed as sexually impure or promiscuous (Betancourt, Borisova, et al., 2008; Burman & McKay, 2007; Coalition to Stop the Use of Child Soldiers, 2008; Mazurana & Carlson, 2006; Mazurana & McKay, 2003; McKay & Mazurana, 2004; McKay, Robinson, Gonsalves, & Worthen, 2006).

Furthermore, prior research has indicated that the length of time that a young person spends with armed groups may place them at greater risk for an accumulation of violence exposures as well as social and material deprivation (Betancourt, Borisova, et al., 2008; Coalition to Stop the Use of Child Soldiers, 2008).

Sources of potential protection have also been documented in research on war-affected youth (Barenbaum et al., 2004; Betancourt & Khan, 2008; Lustig et al., 2004). For child soldiers, acceptance from family members, peers, and others in the community has been indicated as an important postconflict determinant of psychosocial adjustment (Betancourt, Borisova, et al., 2008; Betancourt, Simmons, et al., 2008). For example, Annan et al. (2006) found that male former child soldiers reporting higher levels of family connectedness were likely to have lower levels of emotional distress and better social functioning. In some settings, community members may be open to accepting a child back, despite his or her involvement with armed groups. In other situations, community members may view former child soldiers with fear and distrust due to atrocities committed during war time (Annan et al., 2006; Santacruz & Arana, 2002). In Sierra Leone, despite sensitization campaigns in villages, a widespread initial community response was fear toward the returned child soldiers (Betancourt, Simmons, et al., 2008). Collectively, these studies suggest that the degree to which community acceptance changes over time plays an important ameliorative role for children at risk for poor adjustment given severe forms of war exposure.

Finally, sustained access to education remains an important consideration for the well-being of former child soldiers (Betancourt, Simmons, et al., 2008; Machel, 1996). Opportunities to develop literacy and economic skills can aid in social reintegration of returning children and may act as a deterrent against re-recruitment (Alexander, 2006; Betancourt, Brewer, & De la Soudiere, 2005; Sommers, 2003; Williamson, 2005). Schools may also serve as an important place to build positive social ties and a source of predictability in the lives of war-affected youth (Betancourt, 2005; Sommers, 2003). In this manner, it is possible that children with consistent school access may demonstrate improved psychosocial adjustment.

Study Aims

This study reports on the baseline and follow-up phases of a longitudinal study of former child soldiers in Sierra Leone. Consistent with a stress-accumulation approach, we hypothesized that psychosocial adjustment among former child soldiers would be associated with the degree of exposure to war violence, with the most detrimental effects among groups exposed to severe events (rape and wounding or killing others). In examining hypotheses about stress adjustment, we also hypothesized that the presence of protective factors related to the individual (school access), family (family acceptance), and community (community acceptance) would be associated with improved psychosocial adjustment at follow-up. We anticipated that the presence of protective factors indicative of social acceptance (family and community acceptance) would lead to better psychosocial outcomes, particularly among youth exposed to potentially stigmatizing war experiences (e.g., rape, wounding or killing others).

Method

Participants and Procedures

This investigation was conducted in collaboration with the International Rescue Committee (IRC). Data collection began in 2002 with a baseline assessment (Time 1 [T10]) of 260 former child soldiers who had been affiliated with the RUF in Sierra Leone. At the time, the IRC was a lead agency supporting DDR in several parts of the country. The initial sample for the study was obtained using a two-stage method of selection. First, IRC registries were pooled to create a master list of all youth who had been processed through Interim Care Centers at all sites in four districts of Sierra Leone during the most active period of demobilization, from June 2001 to February 2002 ($n = 309$). All participants were between the ages of 10 and 18 at the time of selection and had been involved with the RUF's fighting forces. Exclusion criteria included inability to be interviewed due to a severe cognitive or physical disability as reported by program social workers (although no such cases were identified) and/or not falling within the target age range.

At follow-up in 2004 (Time 2 [T2]), 56.5% of the original sample was reinterviewed ($n = 156$). Midway through T2 data collection, the death of the IRC's country director and subsequent suspension of program activities halted completion of the follow-up survey. As a result, 31% of the original sample was not recontacted for follow-up. Of those T1 participants attempted for interviews at T2, 1% of caregivers declined consent for their child to participate, 1% of participants had died, and 7% of the initial sample had relocated at a distance that prevented follow-up. Participants lost to follow-up did not differ significantly from ongoing participants in terms of age ($p = .74$), gender ratio ($p = .60$), or literacy levels ($p = .48$). Higher percentages of completers versus noncompleters (71% vs. 53%) attended school at baseline ($p = .002$). Participants lost to follow-up reported similar levels of community acceptance at T1 ($p = .63$). Ongoing participants did not differ from those lost to follow-up on depression ($p = .12$), anxiety ($p = .29$), hostility ($p = .14$), prosocial attitudes ($p = .32$), or confidence ($p = .96$) as assessed at baseline. No significant differences were found between noncompleters and ongoing participants in terms of age of abduction ($p = .19$) or length of time with the fighting forces ($p = .45$).

Data for all participants at T1 and T2 were collected through face-to-face interviews by a team of seven trained Sierra Leonean research assistants, monitored by the study PI and IRC staff at the country level. All research staff participated in a 4-day training on interviewing procedures and ethical practices in working with vulnerable populations. The Sierra Leonean local research team and the study principal investigator determined a process to best introduce the research to the community in a manner respectful of cultural norms about decision making. First, the process of consent involved contact with village leaders and neighborhood meetings to introduce and explain the research. Second, local research assistants traveled to the target households with an IRC staff member. The study was then introduced and any questions answered in the local language, Krio, before a signature was given. All instructions, consents, and assents were read aloud. Participants were informed of the general intent of the study and its longitudinal nature, including the intent to contact participants for future interviews. The youth-informed assent process was conducted separately from the parent- or guardian-informed consent.

Due to a low literacy rate in the study population (50%), all survey protocols were administered verbally. The survey instrument, recruitment materials, and the informed consent forms were constructed for easy comprehension and brevity. Visual aids were used to assist in the administration of Likert-type scales. The T1 assessments were approved by the IRC country program and headquarters staff; the T2 survey received Institutional Review Board approval from Boston University School of Medicine and Boston Medical Center where the first author was based at the time. At baseline and follow-up, IRC social workers traveled with the research team to respond to cases requiring additional attention due to severe emotional or physical health needs.

In non-Western contexts, such as Sierra Leone, the cultural validity of constructs measured remains a perennial challenge for cross-cultural research (Achenbach & Rescorla, 2006; Canino & Alegria, 2008). For these reasons, many of our measurement scales were developed in close consultation with local staff as well as focus group discussions among youth similar to the study population to determine the face validity and cultural relevance of survey items. All measures and questions were then forward- and back-translated to ensure accuracy. Similar combined methods for cross-cultural instrument development have been used in several recent studies of psychosocial adjustment of former child soldiers (Bayer et al., 2007; Kohrt et al., 2008).

This study reports results for the sample of $n = 156$ youth who were contacted at both the T1 and T2 waves of data collection. The average age at T1 in this sample of former child soldiers was 15.13 years, with 12% female and 88% male participants (Table 1). The sample was almost equally split in terms of religious background (53% Christian and 47% Muslim). At T1, 50% of the sample reported poor literacy skills (reading and writing), with similar percentages of both girls and boys reporting poor literacy (53% of girls and 50% of boys); 72% of the youth were attending school at baseline; and 65% were in school at the time of both interviews. At T1, the majority (83%) of the youth reported staying with an immediate family member whereas 17% reported staying with an extended family or other caretaker; 84% of the female and 55% of the male caretakers were illiterate and reported no educational background.

Table 1
Sociodemographic Characteristics of the Sample

Characteristic	*M/N* (*SD*/%)
Age at interview T1, years	15.13 (2.25)
Sex[a]	
Male	137 (88%)
Religion, T1	
Christian	79 (53%)
Muslim	70 (47%)
Literacy, T1	
Poor	73 (50%)
Functional	57 (40%)
Moderate	15 (10%)
In school,[b] T1	107 (72%)
In school at T1 and T2[c]	96 (65%)
Age of abduction, years	10.51 (2.98)
Length of time with armed forces, years	4.68 (0.73)
Relative SES, T2 (range = 4–16)	9.39 (2.28)

Note. *N*s = 143–156.
[a]Coded 0 for boys and 1 for girls. [b]Coded 0 for youth who were not in school and 1 for youth who were in school. [c]Coded 0 for youth who were not in school at T1 or T2, and 1 for youth who were in school at both T1 and T2.

Measures

War experiences. To assess individual exposures to severe forms of war violence, we used items from the Child War Trauma Questionnaire (CWTQ; Macksoud & Aber, 1996), initially developed for use with Lebanese war-affected youth. The questionnaire was adapted by our research team to better capture the context of the war in Sierra Leone; for example, items on bombing and shelling were removed and items on sexual assault and being forced to kill others were added. The adapted instrument contained 42 items regarding child's experience of war-related events. War experiences were coded for occurrence versus no occurrence because some traumas were experienced so commonly that participants had a difficult time accurately recalling or assessing frequency.

Three types of war experiences were examined in light of theory and research on the psychosocial adjustment of former child soldiers: (a) *witnessing general war violence* (e.g., massacres or raids on villages), (b) *experiencing rape*, and (c) *participation in wounding or killing others*. We chose these categories because we were interested in exploring the particular effect of forms of severe violence common in former child soldiers. The survey instrument also included a self-reported age of abduction as well as duration of time with the fighting forces that were analyzed separately.

Retention in school. A variable indicative of sustained access to education was created by comparing youth who reported being in school at both T1 and T2 with youth who were not in school at one or both waves of data collection.

Family acceptance. A measure of family acceptance was drawn from our previous qualitative work and consultations with local staff in 2002. The measure, only administered at T2, consisted of six items assessing the way former child soldiers perceived acceptance, understanding, and respect from their family members. Items were scored on a Likert scale with response options of *not true*, *sometimes true*, or *very true*. Questions on this scale included "since the war you feel you are welcome in the family with whom you live," "you have the same opportunities and responsibilities as other children in the family (or household)," "your caregivers treat you as well as the other children in the household," and so on. This measure exhibited strong internal reliability (Cronbach's α = .93). At T2, reports of family acceptance were skewed toward high values. As a result, the scale was dichotomized for statistical analyses (see next).

Community acceptance. A measure of community acceptance was drawn from our qualitative work and consultations with local staff in 2002. The measure consisted of six items assessing the way former child soldiers perceived acceptance from others in the community. Items were scored on a Likert scale with response options of *not true* to *sometimes true* or *very true*. The measure included items such as "since the war, have people in this community been good to you?" "since the war, you feel you have been welcomed back into the community where you live," and so on. This measure was administered at both waves of data collection and exhibited strong internal reliability (Cronbach's α = .90 at T1 and Cronbach's α = .89 at T2). For each child, we computed a subscale score for community acceptance at T1 and T2. We also computed a "change in acceptance" by subtracting the T1 score from the score at T2. An increase in the community acceptance score between baseline and follow-up would thus indicate a positive change in community acceptance. Because change in community acceptance differed for those starting at higher or lower levels, all analyses were adjusted for T1 community acceptance.

Psychosocial adjustment. To assess psychosocial adjustment, we used a measure developed by researchers at the Oxford Refugee Studies Program (MacMullin & Loughry, 2004). This instrument was drawn from several standardized child mental health measures and was adapted for use among former child soldiers from Sierra Leone and northern Uganda using participatory methods (MacMullin & Loughry, 2004). The measure contains 52 items: three subscales for mental health problems (anxiety, depression, hostility) and two subscales for positive and adaptive adjustment outcomes (confidence and prosocial attitudes). Each of the five domains is represented by a set of questions that inquires about specific behaviors particular to that domain (e.g., "Do you cry easily?" [Depression], "Do you get into fights?" [Hostility], and "Do you share with others?" [Prosocial Attitudes]). For each question responses were scored on a 1–4 scale, ranging from 1 = *never*, 2 = *rarely*, 3 = *sometimes*, and 4 = *always*.

Following our experience with using this measure at T1 in 2002, three items were excluded from the measure when it was readministered at T2 in 2003–2004. Additionally, upon further analyses of T1 and T2 data, three more items were dropped because they did not contribute to the internal consistency of corresponding subscales. The 46-item version of this instrument had good internal consistency across all subscales at both waves of data collection. At follow-up, the Cronbach's α was .67 for the *anxiety subscale* (8 items), .70 for the *depression subscale* (8 items), .88 for the *hostility subscale* (12 items), .76 for the *prosocial attitudes subscale* (10 items), and .69 for the *confidence subscale* (8 items). For each child, we computed a subscale score (e.g., anxiety, depression, hostility, prosocial attitudes, and confidence) at T1 and T2.

Participant demographic characteristics. A demographic inventory collected information on gender, age, and family socioeconomic resources. The measure of socioeconomic status was created to indicate an individual's access to food, housing, and clothing relative to others in the community. The measure consisted of four items and had a good internal consistency (α = .78).

Statistical Analyses

Statistical analysis of the data proceeded in three stages. First, descriptive statistics of study variables were analyzed. Summary statistics for continuous measures are reported as mean and standard deviation, whereas dichotomous measures are reported as percentages along with their associated counts.

Multiple imputation was then used to generate five complete datasets for the 156 participants in the baseline and follow-up interviews (Rubin, 1987). All remaining analyses were conducted using the multiply imputed data sets, and results were combined across imputations. This approach reduces bias to the extent that values on the observed variables are informative about values on items that were missing, increases precision relative to a complete case analysis with a smaller N, and accounts for the sampling variability across imputations (Rubin, 1987). Imputed data sets were generated using the method of chained equations as implemented in *IVEware* (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001) and the results of analyses conducted across imputations were combined using the MIANALYZE procedure using the SAS software system, version 9.1 (SAS Institute, 2004).

First, paired *t* tests were used to compare values for psychosocial adjustment at T1 compared to T2. Second, correlation coefficients were calculated between all variables in order to inspect unadjusted, bivariate associations. Third, multiple linear regression was used to assess the effect of predictors on T2 psychosocial outcomes after adjusting for T1 scores and potential confounders. For each psychosocial outcome (depression, anxiety, hostility, confidence, and prosocial attitudes), three main-effects regression models were fit, allowing assessment of the impact of introducing successive sets of covariates. The first model included only demographic factors (age, sex of child, socioeconomic status [SES]) and baseline psychosocial adjustment scores. In the second model, dichotomous war experiences were added to the model to test the effects of stress accumulation. Model 3 added protective factors to the model to test hypotheses regarding stress adjustment. At each step, we compared parameter estimates and model fits and calculated the difference in R^2.

After assessing the main effects models, we tested whether the inclusion of interaction terms improved the overall model fit. Interactions were examined in conceptual groups testing first interactions between war exposures and protective factors. A second group of interactions was then tested examining interactions between gender and risk factors and then gender and protective factors on psychosocial outcomes. An *F* test using the TEST command in PROC MIANALYZE (SAS Institute, 2004) was performed upon adding each conceptual group of interactions to the final main effects model (Model 3 for each outcome) to test the joint hypothesis that the combined effect of the interactions did not differ from zero. If the *F* test was nonsignificant, no further individual testing of interactions was conducted to reduce risk of chance findings. If the *F* test indicated that interactions in the conceptual group improved the overall model fit, individual interactions were tested by adding them one by one to the final model for each outcome.

All statistical analyses (two-tailed) were performed using the SAS software system, version 9.1 (SAS Institute, 2004).

Results

Exposure to Violence of Study Population

The average age at abduction in this sample of former child soldiers was 10.51 years with no significant difference between males and females. Twelve percent of participants reported joining the fighting forces at very young ages (2–5 years), which likely reflects the fact that some children may have been abducted while with their mothers in the bush. Average length of abduction was 4.68 years with males and females reporting a similar duration of time in captivity with armed forces (Table 1). Nearly all (99%) former child soldiers in the sample reported forced conscription (mainly through abduction).

All former child soldiers in the sample were exposed to high levels of violence. For example, 89% of females and 88% of males had witnessed some sort of war-related violence such as massacres or raids on villages. Being a victim of rape was reported by 37% of females and 6% of males in the sample. Perpetration of extreme acts of violence was reported by both males and females; 26% of females and 28% of males reported having killed either a stranger or loved one. Overall, 7% of youth reported experiencing all three exposures—witnessing mass violence, rape, and wounding or killing others.

Change in Psychosocial Outcomes

Table 2 reports the results of the baseline and follow-up assessments of each of the subscales on the measure of psychosocial adjustment. Average scores for the depression, hostility and anxiety subscales did not differ significantly between T1 and T2. In terms of the adaptive outcomes, confidence scores were similar between the two waves of assessment, yet average prosocial attitudes were lower at T2 compared with T1.

Predictors of Psychosocial Outcomes: Correlational Analysis

Bivariate correlations indicated expected patterns of covariation within and between variable domains (Table 3). The three subscales of mental health problems were positively intercorrelated as were the subscales for the two adaptive outcomes, confidence and prosocial attitudes ($r = .73$). Higher levels of prosocial attitudes and confidence at T2 were correlated with symptoms of anxiety ($r = .22$, and $r = .27$, respectively), but not other mental health problems.

Table 2
Psychosocial Adjustment Over Time

	Range	T1, *M* (*SD*)	T2, *M* (*SD*)	Change over time, *M* (*SD*)	*p* value[a]
Anxiety	8–32	19.68 (5.62)	20.23 (4.37)	0.55 (6.25)	.27
Depression	7–28	14.77 (3.62)	14.97 (3.87)	0.19 (4.62)	.61
Hostility	13–52	18.93 (4.87)	19.91 (6.12)	0.98 (7.24)	.09
Confidence	8–32	24.65 (4.37)	24.78 (4.12)	0.13 (5.12)	.74
Prosocial attitudes	10–40	34.99 (4.12)	34.01 (4.12)	−0.98 (5.37)	.02

Note. *N* = 156.
[a]Paired *t* tests.

Table 3
Correlation Between Variables

Outcome variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Demographics																
1. Age	—															
2. Female	.04	—														
3. Socioeconomic status	−.05	.13	—													
War exposure																
4. Age of abduction (T1)	.63***	−.01	−.15†	—												
5. Length of time with armed forces (T1)	−.04	−.06	.01	−.40***	—											
6. Wounding or killing others	.15†	0	−.01	−.06	.08	—										
7. Victim of rape	−.02	.62***	−.07***	−.08	.04	.64***	—									
8. Witness to violence	−.03	.05	.02	0	−.02	.48**	.18**	—								
Protective factors																
9. Retention in school (T1, T2)	−.09	−.21	−.02	−.01	−.05	.01	−.39*	−.16	—							
10. Family acceptance	−.28***	−.25***	.11	−.04	−.08	−.27*	−.28***	−.29***	.24†	—						
11. Community acceptance over time (T1, T2)	−.01	.04	.21**	−.05	.11	−.11	.01	−.09	−.09	.34***	—					
Outcomes																
12. Depression symptoms	.14†	.18*	−.03	−.1	.03	.24**	.24**	.14†	−.1	−.33***	−.15†	—				
13. anxiety symptoms	.09	.08	−.06	−.1	.08	.31***	.38***	.11	−.06	−.22**	−.05	.67***	—			
14. Hostility symptoms	.05	.21*	−.03	−.09	.06	.40***	.49***	.20*	−.08	−.30***	−.16*	.58***	.62***	—		
15. Prosocial attitudes	−.15†	−.12	.22**	−.14†	.13	−.02	.21***	−.05***	.15†	.26**	.22**	−.09	.22**	.01	—	
16. Confidence	−.07	−.16*	.24**	−.11	.07	.01	.08	.03	.11	.23**	.11	.09	.27***	.02	.73***	—

Note. *N* = 156. Correlations between dichotomous variables and continuous outcomes are assessed using Spearman's rank sum correlation and between two binary variables assessed using polychoric correlations.
$^{\dagger}p \leq .10$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Turning to cross-domain relations, we explored the relation between the adjustment outcomes and two sets of predictors: (a) war-related risk factors (endorsing general witnessing of war violence, experiencing rape, wounding/killing others, length of time with the RUF, and age of abduction) and (b) hypothesized protective factors (retention in school, community acceptance, family acceptance). Several of these variables were associated with T2 psychosocial outcomes.

War-related risk factors. Witnessing of general war-related violence was correlated with higher levels of hostility at T2 (*r* = .20). Being abducted at a younger age and length of time spent with the fighting forces were not significantly correlated with any of the outcomes investigated. Experiencing rape during the war was associated with higher levels of hostility at T2 (*r* = .49). In addition, rape was associated with higher levels of depression (*r* = .24) and anxiety (*r* = .38) at T2. Rape was also associated with higher levels of prosocial behaviors (*r* = .21) yet inversely correlated with staying in school (*r* = −.39) and family acceptance (*r* = −.28). Having killed someone during the war was associated with higher levels of depression ($r = .24$), anxiety ($r = .31$), and hostility ($r = .40$).

Protective factors. Of the protective factors examined, family and community acceptance demonstrated promise as important psychosocial resources for the youth in our sample. Family acceptance was inversely correlated with depression ($r = -.33$), anxiety ($r = -.22$), and hostility ($r = -.30$), and positively correlated with prosocial attitudes ($r = .26$) and confidence ($r = .23$) at T2. Community acceptance was inversely correlated with hostility ($r = -.16$) and positively correlated with prosocial attitudes ($r = .22$) at T2. Retention in school was only marginally correlated with higher levels of prosocial attitudes and behaviors ($r = .15$).

Predictors of Psychosocial Outcomes Over Time: Multiple Regression Analysis

Three potentially confounding variables were examined to determine whether they were associated with the variables of interest in this study: gender, age of child, and relative SES. Females in our sample were significantly more likely to experience rape ($p \leq .0001$). Our analyses also revealed that females had higher levels of hostility at T2 ($M = 22.67$, $SD = 4.67$) compared to males in the sample ($M = 19.47$, $SD = 6.16$, $t = 2.12$, $p \leq .05$). Females and males reported comparable levels of depression and anxiety symptoms as well as confidence. Age of participants was correlated with age of abduction (with older participants in the sample being abducted at older ages). Older youth had significantly lower levels of family acceptance ($r = -.28$, $p \leq .001$). Age was not significantly correlated with any other psychosocial outcomes investigated. Socioeconomic status was correlated with higher community acceptance at T2 ($r = .19$, $p \leq .05$). Higher SES was also correlated with higher average levels of prosocial attitudes ($r = .22$, $p \leq .01$) and confidence ($r = .25$, $p \leq .01$). Therefore, gender, age, and SES were entered as covariates in all regression analyses.

War-related violence and mental health adjustment. Table 4 displays regression analyses exploring levels of depression, anxiety, hostility, prosocial attitudes, and confidence at follow-up (T2). Adjusting for demographic variables and T1 scores, youth abducted into the fighting forces when they were younger ($\beta = -0.41$, $p \leq .05$) were likely to report more symptoms of depression at T2. In examining the effects of stress accumulation, we observed that witnessing general violence, experiencing rape, and having killed others during the war were not predictive of depression symptoms at T2, adjusting for all covariates.

None of the demographic variables investigated had an effect on anxiety level at T2. Witnessing general war-related violence was not a significant predictor of anxiety at follow-up while being involved in wounding or killing others was only marginally significant in predicting symptoms of anxiety ($\beta = 1.58$, $p \leq .10$). Of all risk factors studied, surviving a rape was the strongest predictor of anxiety at T2 controlling for anxiety levels at T1 ($\beta = 4.06$, $p \leq .05$) and adjusting for all other covariates. A history of having survived a rape contributed to an anxiety score more than half a standard deviation greater at follow-up.

In terms of hostility, surviving a rape predicted significantly higher levels of hostility at T2 ($\beta = 7.42$, $p \leq .001$) such that surviving a rape contributed to a hostility score a full standard deviation greater at follow-up upon adjusting for all other factors. A history of wounding or killing others ($\beta = 3.06$, $p \leq .05$) was also associated with hostility scores nearly a half standard deviation higher than the sample mean at follow-up. These patterns were consistent when the protective factors were added to the model in Step 3. Length of time with armed groups and age of abduction were not significantly associated with levels of hostility at follow-up (T2).

For adaptive outcomes, our analysis revealed that gender and SES were both strong predictors of prosocial attitudes and confidence levels at follow-up. In particular, females had significantly lower prosocial attitudes ($\beta = -3.13$, $p \leq .01$) and confidence ($\beta = -2.85$, $p \leq .05$) at T2, whereas living in wealthier families was associated with increased levels of prosocial attitudes ($\beta = 0.26$, $p \leq .05$) and confidence ($\beta = 0.34$, $p \leq .01$). Neither general witnessing of violence nor a past history of wounding or killing others played a significant role in predicting levels of confidence or prosocial attitudes. However, survivors of rape demonstrated higher levels of both prosocial behaviors ($\beta = 5.22$, $p \leq .001$) and confidence at follow-up ($\beta = 3.40$, $p \leq .05$).

Individual, family, and community factors and psychosocial adjustment. Among the protective factors examined, staying in school and changes in community acceptance exerted the largest positive impact on psychosocial adjustment, indicating their importance for facilitating stress adjustment in this setting. In particular, increases in community acceptance between T1 and T2 (adjusting for baseline community acceptance) were significantly associated with lower depression ($\beta = -0.35$,

Table 4

Multiple Regression Analyses for Psychosocial Outcomes

	Depression				Anxiety				Hostility				Prosocial attitudes				Confidence			
Variables	R^2 ΔR^2	Model 1	Model 2	Model 3	R^2 ΔR^2	Model 1	Model 2	Model 3	R^2 ΔR^2	Model 1	Model 2	Model 3	R^2 ΔR^2	Model 1	Model 2	Model 3	R^2 ΔR^2	Model 1	Model 2	Model 3
Step 1																				
Covariates	.08				.07				.07				.11				.17			
Baseline score		.19* (0–0.38)	.18† (0–0.36)	.14 (−0.06–0.33)		.18* (0.05–0.31)	.16* (0.03 to 0.28)	.17* (0.04 to 0.29)		.19† (−0.01 to 0.39)	.07 (−0.11 to 0.25)	.08 (−0.10 to 0.27)		.13 (−0.03 to 0.29)	.14† (−0.01 to 0.30)	.09 (−0.06 to 0.24)		.26** (0.11 to 0.41)	.30** (0.14 to 0.45)	.24* (0.08 to 0.39)
Age (T1)		.19 (−0.08 to 0.46)	.52* (0.13 to 0.90)	.36† (−0.04 to 0.76)		.13 (−0.18 to 0.43)	.38† (−0.02 to 0.79)	.31 (−0.12 to 0.75)		.13 (−0.29 to 0.56)	.16 (−0.38 to 0.69)	−.02 (−0.57 to 0.54)		−.25† (−0.53 to 0.03)	−.23 (−0.61 to 0.15)	.03 (−0.34–0.39)		−.14 (−0.41 to 0.12)	−.03 (−0.40 to 0.35)	.17 (−0.23 to 0.56)
Female		1.57† (−0.29 to 3.43)	.92 (−1.01 to 2.85)	.69 (−1.20 to 2.57)		.72 (−1.34 to 2.78)	−.49 (−2.52 to 1.53)	−.78 (−2.85 to 1.29)		3.80* (1.00 to 6.60)	1.91 (−0.70 to 4.53)	1.39 (−1.21 to 4.00)		−2.02* (−3.88 to −0.17)	−3.46** (−5.38 to −1.54)	−3.13** (−4.89 to −1.38)		−2.06* (−3.85 to −0.28)	−3.12* (−5.02 to −1.23)	−2.85* (−4.72 to −0.98)
Relative SES (T2)		−.09 (−0.36 to 0.17)	−.14 (−0.39 to 0.12)	.02 (−0.25 to 0.29)		−.10 (−0.40 to 0.20)	−.11 (−0.39 to 0.17)	−.08 (−0.38 to 0.22)		−.15 (−0.57 to 0.27)	−.10 (−0.47 to 0.27)	.07 (−0.31 to 0.45)		.42* (0.14 to 0.69)	.48** (0.21 to 0.75)	.26* (0 to 0.52)		.46** (0.19 to 0.72)	.48** (0.22 to 0.75)	.34* (0.07 to 0.62)
Step 2	.20				.25				.35				.23				.23			
War exposures	.12				.18				.28				.12				.06			
Age of abduction			−.41* (−0.74 to −0.07)	−.31† (−0.64 to 0.02)			−.33† (−0.67 to 0.02)	−.30† (−0.65 to 0.05)			−.12 (−0.57 to 0.33)	−.03 (−0.47 to 0.42)			.05 (−0.26 to 0.35)	−.10 (−0.39 to 0.19)			−.09 (−0.39 to 0.22)	−.19 (−0.50 to 0.12)
Length of time with RUF			−.35 (−1.32 to 0.63)	−.25 (−1.22 to 0.71)			.08 (−0.96 to 1.11)	.05 (−0.99 to 1.10)			.30 (−1.07 to 1.67)	.42 (−0.93 to 1.78)			.65 (−0.30 to 1.59)	.48 (−0.39 to 1.34)			.22 (−0.71 to 1.14)	.13 (−0.78 to 1.05)
Witnessing			1.18 (−0.61 to 2.96)	.78 (−0.98 to 2.53)			.59 (−1.34 to 2.52)	.50 (−1.46 to 2.47)			2.41† (−0.12 to 4.94)	2.02 (−0.50 to 4.53)			−.36 (−2.20 to 1.49)	.39 (−1.29 to 2.07)			.74 (−1.08 to 2.56)	1.15 (−0.64 to 2.94)
Victim of sexual abuse			1.53 (−0.75 to 3.80)	1.74 (−0.53 to 4.00)			4.02* (1.54 to 6.50)	4.06* (1.49 to 6.62)			7.09*** (3.85 to 10.33)	7.42*** (4.17 to 10.68)			5.15*** (2.78 to 7.52)	5.22*** (3.03 to 7.40)			3.57* (1.21 to 5.93)	3.40* (1.06 to 5.74)
Wounding or killing others			1.12 (−0.39 to 2.63)	.82 (−0.67 to 2.31)			1.67* (0.05 to 3.29)	1.58† (−0.08 to 3.24)			3.49* (1.36 to 5.62)	3.06* (0.92 to 5.19)			−1.08 (−2.66 to 0.49)	−.85 (−2.28 to 0.59)			−1.09 (−2.66 to 0.47)	−.80 (−2.34 to 0.74)
Step 3	.27				.27				.39				.39				.29			
Protective factors	.07				.02				.04				.16				.06			
Retention in school				−.08 (−1.28 to 1.11)				.40 (−0.94 to 1.73)				.64 (−1.08 to 2.37)				1.55* (0.37 to 2.72)				.56 (−0.71 to 1.82)
Family acceptance				−.89 (−2.59 to 0.81)				−1.41 (−3.26 to 0.45)				−2.17† (−4.55 to 0.21)				.17 (−1.43 to 1.76)				.58 (−1.13 to 2.29)
Δ in community acceptance				−.35* (−0.64 to −0.06)				.05 (−0.27 to 0.37)				−.27 (−0.68 to 0.14)				.62*** (0.35 to 0.90)				.32* (.03 to 0.62)

Note. $N = 156$. For each predictor, we report parameter estimate, associated p value and 95% confidence intervals. R^2 = percentage of outcome variance explained by each regression model; ΔR^2 = difference in percentage of outcome variance explained between regression models; SES = socioeconomic status.

$^{\dagger}p \leq 0$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

$p \leq .05$), increased prosocial attitudes and behaviors ($\beta = 0.62$, $p \leq .001$), and increased confidence ($\beta = 0.32$, $p \leq .05$) at T2. Staying in school between T1 and T2 was predictive of greater prosocial attitudes at follow-up ($\beta = 1.55$, $p \leq .05$). Although it did not reach the $p \leq .05$ level of significance, perhaps in part due to 77% of subjects giving the highest possible rating on this scale, increased family acceptance demonstrated an inverse relation with hostility ($\beta = -2.17$, $p \leq .10$). None of the other protective covariates appeared to exert a positive effect on anxiety. None of the conceptual groups of interaction terms were found to significantly improve the fit of these final main effects models as assessed by joint *F* tests.

Discussion

In this analysis, we obtained partial support for both a risk accumulation and a stress adjustment model of psychosocial well-being among former child soldiers. We tested the association between exposure to war-related experiences of violence and psychosocial adjustment over a 2-year period of follow-up. In addition, we were interested in the role of protective factors at the individual, family, and community levels in terms of their relation with improved adjustment. Finally, we explored the potential mitigating impact of significant individual, family, and community protective factors in the relation between war experiences and psychosocial adjustment. Although average levels of depression, anxiety, and hostility remained stable between baseline and follow-up, important changes were observed among a number of subgroups in the sample.

War-Related Experiences and Mental Health Adjustment

Although most children who survive wars are exposed to some degree of violence, our analyses indicated that two forms of violence common among former child soldiers—wounding or killing others and rape—had a particularly toxic influence on long-term psychosocial adjustment, particularly with regard to the outcomes of hostility and anxiety. These two categories of war traumas are set apart from the widespread witnessing of violence commonly reported among war-affected youth (Barenbaum et al., 2004; Lustig et al., 2004) by their intimate nature and intensity. However, research on child soldiers to date has not examined the specific longitudinal influence of these more severe forms of violence exposure.

Standard approaches to the measurement of war-related violence typically add all war-related experiences into a single total exposure score. Such an approach may mask important dynamics related to the most severe forms of violence exposure, particularly those that distinguish the experiences of former child soldiers (Hatch & Dohrenwend, 2007; Hundt, Chatty, Thabet, & Abuateya, 2004; Netland, 2005). Our work builds on the work of Macksoud and Aber (1996) who developed the CWTQ to address the need to make qualitative distinctions between different types of war-related experiences faced by children and youth. Given the widespread trauma, violence, and loss that characterize most war-affected young people, it is important to distinguish between forms of violence that are frightening—but widespread—and those that are less common—but distinctly toxic—in order to accurately identify those at risk for poor adjustment over time. It is important to note that witnessing of general war violence, although very common in this sample, was not a strong predictor of psychosocial adjustment over time. In contrast, the effects of experiencing rape and wounding or killing others were longlasting.

Of all the psychosocial outcomes investigated, the significant increases in hostility observed among youth who reported rape or wounding or killing are particularly concerning in that these exposures are much more likely to accumulate in child soldiers. Participation in wounding or killing others was a significant predictor of hostility and remained a marginal predictor of anxiety even upon adjusting for all other variables examined in these analyses. The relations we observed between extreme war experiences and these mental health problems are consistent with observations made in other studies of former child soldiers (Annan et al., 2006; Santacruz & Arana, 2002) as well studies of adult war veterans (Hofmann, Litz, & Weathers, 2003; Maguen et al., 2006) and adult combatants in neighboring Liberia (Johnson et al., 2008).

Although the present analysis was limited by a small sample of females, gender was an important predictor of decreased prosocial attitudes and lower confidence over time. This finding likely indicates the ''double indemnity'' facing female former child soldiers in this context who were not only involved in violence perpetration but also frequently face greater stigma upon return home. Our qualitative data have indicated that these young women face increased risk of community stigma due to perceptions that they have been made impure or promiscuous by sexual violation (Betancourt, Simmons, et al., 2008).

The finding in our data of an association between surviving a rape and increased confidence and prosocial attitudes was unexpected but must be considered in light of the dynamics of the Sierra Leone conflict. It is important to note that, on average, the youth in our sample had spent more than 5 years with the RUF. The RUF were well known for their brutality and harsh techniques of indoctrination and suppression, particularly in their treatment of child soldiers. Given this context, any former child soldier who made it through the war alive likely developed and relied upon a number of survival strategies to navigate such a harsh and dangerous environment. Some of these young people, particularly those who survived abuse, may actually possess a sense of resourcefulness that is exhibited in improved confidence and sense of agency over time. Indeed, many of the items measured in the confidence subscale capture such dynamics (e.g., feeling calm under pressure, confident in doing things on one's own).

In our qualitative data, prosocial behaviors, particularly the efforts of former child soldiers to be kind to others and have positive interactions, were described as a means of signaling to others that a young person was not a threat or morally corrupt. As a female former child soldier explained: "My parents accepted me … but other people in the community [did] not understand. … I later mingled with them [after that they] have observed that I am not a bad person. … I was very kind to everybody and we started to do things in common."

Our finding of a significant relation between a younger age of abduction and increased levels of depression over time aligns with recent developmental science indicating that exposure to toxic stress may be most damaging during sensitive periods in a child's development (National Scientific Council on the Developing Child, 2005; Shonkoff & Phillips, 2000). In addition, the lack of association between length of time with the fighting forces and any of the psychosocial outcomes (adjusted for severe violence exposures) indicates that severe violence exposures, rather than absolute length of abduction, are most important for predicting psychosocial adjustment over time.

Protective Factors and Mental Health Adjustment

Just as the presence of distinct toxic exposures must be considered in identifying which war-affected youth are most at risk for poor psychosocial adjustment, the presence of certain protective processes must also be evaluated. Consistent with research on the psychosocial benefits of school participation for children affected by armed conflict (Betancourt, 2005; Elbedour, Bensel, & Bastien, 1993; Loughry et al., 2006; Sommers, 2003), we found that staying in school was associated with improved prosocial attitudes and behaviors. In the Sierra Leonean context, failure to find a greater protective effect of school retention on negative mental health outcomes such as anxiety, depression, and hostility may be due to poor-quality schools and the limited availability of psychosocial supports in the present education system. Despite the fact that Sierra Leone's secondary schools serve a cohort of youth whose lives were deeply affected by the war, very few secondary schools offer teacher salaries that lead to quality teaching, let alone time, training, or staff to provide support and guidance for traumatized children (Betancourt, Simmons, et al., 2008).

The Mitigating Effect of Community Acceptance

Overall, community acceptance emerged from our analysis as an important predictor of reduced depression over time as well as improvements in prosocial attitudes and confidence. These findings are consistent with other research pointing to community acceptance as an important factor in understanding the long-term reintegration of former child soldiers (Allen & Schomerus, 2006; Boyden & de Berry, 2004; Bracken & Petty, 1998). Importantly, the influence of community acceptance was found to have a protective effect (Luthar, Cicchetti, & Becker, 2000) on depression and an enhancing effect on confidence and prosocial attitudes of former child soldiers regardless of whether they had a history of the most severe violence exposures. This finding speaks to the importance of postconflict efforts to sensitize communities and encourage community acceptance of former child soldiers. In return, it appears that when former child soldiers experience improvements in community acceptance, they may return the gesture with increased positive effort (i.e., helping others, sharing, initiating activities). This finding speaks further to the agency of war-affected youth who, despite exposure to some of the most horrendous experiences of war, may make an extra effort to be helpful and kind to others when the larger community projects a willingness to accept them.

Overall, our study identified only a few protective factors that might assist the recovery of former child soldiers who have experienced the most severe and cumulative forms of war trauma. In fact, the magnitude of the effects of war-related stressors, particularly surviving rape and being involved in wounding or killing, were much larger than those associated with protective factors such as staying in school and experiencing improvements in community acceptance. Our finding that the stress-adjustment effects in this sample are generally weaker than the stress-accumulation effects has important public health implications. It suggests that we have a long way to go before being able to fully mitigate the effects of particularly toxic stressors in the lives of war-affected youth, such as rape and involvement in wounding or killing others.

In future research, the search for additional sources of protection must be broadened to consider other forms of protection and support. It also remains important to evaluate the influence of access to traditional healing resources as well as more formal clinical interventions in this population. Despite the severe stressors characterizing some former child soldiers, it is important to note that there has yet to be a rigorous evaluation of interventions addressing the influence of such traumatic events (or postconflict aftereffects such as stigma) on child mental health and development.

Overall, our findings indicate no justification for an approach to targeting services at war-affected youth based on simple labels such as *former child soldiers* without individualized assessment. Although wounding or killing others and surviving rape were important risk factors in our sample, these exposures did not characterize the majority of the sample. However, for youth with accumulated exposures to the most toxic forms of violence, access to rehabilitative services, such as clinical mental health treatment or other traditional approaches to healing, may be some of the few pathways forward.

Further longitudinal research is necessary to truly disentangle the processes involved in the psychosocial adjustment of former child soldiers, the effects of extreme war exposures, and their relation to protective processes over time. In future research, more information is needed about the functions of family and peer relations as resources and supports, as well as potential sources of continued strain in the process of reintegration. Beyond retention in school, the degree to which former child soldiers have been involved in other skills training programs, income-generating activities, and the degree to which they have access to other mentors or positive peers must be examined.

Limitations

There are a number of study limitations must be noted. First of all, an inherent challenge within cross-cultural research is studying constructs in non-Western settings, where local notions of family or community acceptance are likely to be shaped by local context (Canino & Alegria, 2008; Honwana, 2006; Matias-Carrelo et al., 2003). To ensure cultural relevance, we addressed these challenges through gathering additional qualitative data about the constructs of interest as well as consulting local staff members and others working with war-affected youth. Although many of the constructs we studied, such as family and community acceptance, have particular relevance to Sierra Leone, to a large degree these constructs mirror mainstream theory about indicators of acceptance that have been used in prior research on the reintegration of adult soldiers following conflicts such as Vietnam (King, King, Fairbank, Keane, & Adams, 1998).

Further, the sensitive nature of the data collected may be affected by response bias, particularly for difficult events such as rape and perpetration of wounding or killing. Previous research with survivors of violence has suggested that errors in reporting violent experiences, particularly in underreporting, may derive from deliberate strategies on the part of study participants to keep violent events secret in an effort to avoid retribution, future harm, or stigma (Guterman & Cameron, 1997; Guterman, Cameron, & Staller, 2000). It is possible that this is the case in our sample, especially for the reporting of rape and perpetration, which would likely result in underestimations of the relations examined in this study.

In addition, age of the child and the time between the violent event and the time of reporting are also important factors to consider (Brandt, Ward, Dawes, & Flisher, 2005; Guterman et al., 2000; Tajima, Herrenkohi, Huang, & Whitney, 2004). In general, younger children (preschool age) tend to have a harder time accurately reporting violent events that may have happened to them (Tajima et al., 2004). However, in our data, these concerns are reduced somewhat by the fact that the respondents were adolescents reporting on experiences with violence only 2 years after the end of the war; thus, their reporting bias may be more akin to adult responses in similar situations.

The lack of baseline data on prewar mental health problems among youth in Sierra Leone limits our ability to compare the indicators of psychosocial adjustment observed in this sample to the general population. In addition, the challenge of identifying and retaining an appropriate comparison group remains one of the most persistent challenges in conducting research of this nature. Due to the dynamics of the war, it is impossible to identify a group of youth comparable with this cohort who did not experience some degree of exposure to war-related violence. Although our study design initially included a comparison group, the loss to follow-up in this group between T1 and T2 was too great to allow for reliable comparisons, a challenge that we intend to remediate in future stages of this research.

Finally, because our sample represents a cohort of youth demobilized over a specified time period rather than a regionally representative sample, our findings are limited in their ability to be generalized nationally. However, because our sample draws from more than four districts in Sierra Leone, we have no reason to believe that findings from this sample are not reflective of dynamics observed in other parts of the country.

Implications

It is important to consider the findings presented here in light of the current situation of social and mental health services for war-affected youth in Sierra Leone. At present, policy makers in Sierra Leone and the international community are focused on economic development and stability for the country. However, failure to address the consequences of the war on Sierra Leone's young generation threatens the development of human capital and the investments being made in education and youth employment schemes (Mac-Ikemenjima, 2008). Our finding of a progressive deterioration over time in the adjustment of the most highly traumatized former child soldiers, contributing to increased levels of hostility for those who had killed, and increased anxiety and hostility among rape survivors, is of particular concern. Although rape survivors demonstrate more complexity in their response to war (not only manifest in greater risk for mental distress but also a capacity for improved confidence and prosocial attitudes), those with a history of wounding or killing others show only negative outcomes with time. In Sierra Leone, it is important to note that a number of programs have evolved to attend to the consequences of rape and other forms of gender-based violence; however, very little attention has been given to addressing the mental health consequences of perpetrating atrocities. Certainly, research from other wars has indicated that such experiences have profound developmental consequences and pose persistent risks for mental health problems (Davison et al., 2006; Grossman, 1996; King & King, 1999).

Without more targeted attention to support these young people to make the best of the opportunities available to them in development efforts, the risks facing this particularly vulnerable cohort are likely to become more intractable. Such lost opportunity will have personal consequences for the youth in question, a compelling enough reason to act on human rights grounds alone. However, failure to truly address such long-term challenges to true reintegration and *rehabilitation* could very well have broader sociopolitical and security consequences for the society and country at large.

Unfortunately, many DDR programs operating internationally are hampered by inadequate funds to address the social services needs of war-affected youth and families over the long term. The funding streams, technical assistance, and human resources developed during the initial humanitarian emergency and DDR period have not translated into sustainable systems of care. Although significant efforts were made in Sierra Leone's DDR efforts to provide assistance to former child soldiers in the immediate postconflict period, very few social or mental health services for war-affected youth of any kind remain today. With good reason, recent policy efforts have been cautious to rely on labeling or the targeting of only certain groups (such as former child soldiers) for services. However, in the attempt to avoid a focus on psychopathology or the singling out of specific groups for care in isolation of others, the reality of the residual trauma that some of these youth face has been minimized. As our data indicate, despite the dangers of labeling, there is an alarming concentration of particularly toxic risk factors among a portion of former child soldiers. Failing to ensure long-term monitoring, assessment, and follow-up of these particularly high-risk groups avoids the issue.

In conclusion, the present study contributes to the scant literature on the experiences of child soldiers globally by providing the first prospective longitudinal data on a cohort of males and females. It is our hope that this and future research can enhance our scientific understanding of the processes linking war-related traumas to psychological functioning and raise awareness among local governments and the international community of the importance of investing in reasonable, effective, and sustainable responses to support the mental health needs of *all* war-affected children.

References

Achenbach, T. M., & Rescorla, L. A. (2006). *Multicultural understanding of child and adolescent psychopathology: Implications for mental health assessment*. New York: Guilford.

Alexander, J. (2006). *Community based reintegration: Programme evaluation*. New York: UNICEF.

Allen, T., & Schomerus, M. (2006). *A hard homecoming: Lessons learned from the reception center process in Northern Uganda*. Washington, DC: Management Systems International.

Annan, J., Blattman, C., Carlson, K., & Mazurana, D. (2008). *The state of female youth in northern Uganda: Findings from the survey of war affected youth*. Uganda: UNICEF.

Annan, J., Blattman, C., & Horton, R. (2006). *The state of youth and youth protection in Northern Uganda*. Uganda: UNICEF.

Ashby, P. (2002). Child combatants: A soldier's perspective. *Lancet, 360*(9350), s11.

Barenbaum, J., Ruchkin, V., & Schwab-Stone, M. (2004). The psychosocial aspects of children exposed to war: Practice and policy initiatives. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45*, 41–62.

Bayer, C. P., Klasen, F., & Adam, H. (2007). Association of trauma and PTSD symptoms with openness to reconciliation and feelings of revenge among former Ugandan and Congolese child soldiers. *JAMA, 298*, 555–559.

Beckett, C., Maughan, B., Rutter, M., Castle, J., Colvert, E., Groothues, C., et al. (2006). Do the effects of early severe deprivation on cognition persist into early adolescence? Findings from the English and Romanian adoptees study. *Child Development, 77*, 696–711.

Betancourt, T. S. (2005). Stressors, supports and the social ecology of displacement: Psychosocial dimensions of an emergency education program for Chechen adolescents displaced in Ingushetia, Russia. *Culture, Medicine and Psychiatry, 29*, 309–340.

Betancourt, T. S., Borisova, I., Rubin-Smith, J., Gingerich, T., Williams, T., & Agnew-Blais, J. (2008). *Psychosocial adjustment and social reintegration of children associated with armed forces and armed groups: The state of the field and future directions*. Austin, TX: Psychology Beyond Borders.

Betancourt, T. S., Brewer, S., & De la Soudiere, M. (2005). *Holding on to hope: Education, coping, and the struggles of children associated with fighting forces in Sierra Leone. A report to the international rescue committee children and youth protection and development program*. New York: International Rescue Committee.

Betancourt, T. S., & Khan, K. T. (2008). The mental health of children affected by armed conflict: Protective processes and pathways to resilience. *International Review of Psychiatry, 20*, 317–328.

Betancourt, T. S., Pochan, S., & de la Soudiere, M. (2005). *Psychosocial adjustment and social reintegration of child ex-soldiers in Sierra Leone: A follow-up analysis*. Freetown: International Rescue Committee.

Betancourt, T. S., Simmons, S., Borisova, I., Brewer, S. E., Iweala, U., & de la Soudiere, M. (2008). High hopes, grim reality: Reintegration and the education of former child soldiers in Sierra Leone. *Comparative Education Review, 52*, 565–587.

Blattman, C., & Annan, J. (2007). *The consequences of child soldiering*. Brighton: The Institute of Development Studies, University of Sussex.

Boothby, N. (2006). When former child soldiers grow up: The keys to reintegration and reconciliation. In N. Boothby, A. Strang, & M. Wessells (Eds.), *A world turned upside down: social ecological approaches to children in war zones* (pp. 155–178). Bloomfield, CT: Kumarian Press.

Boothby, N., Crawford, J., & Halperin, J. (2006). Mozambique child soldier life outcome study: Lessons learned in rehabilitation and reintegration efforts. *Global Public Health, 1*, 87–107.

Boyden, J., & de Berry, J. (Eds.). (2004). *Children and youth on the front line: Ethnography, armed conflict and displacement*. New York: Berghahn Books.

Bracken, P., & Petty, C. (Eds.). (1998). *Rethinking the trauma of war*. London: Free Association Books.

Brandt, R., Ward, C. L., Dawes, A., & Flisher, A. J. (2005). Epidemiological measurement of children's and adolescents' exposure to community violence: Working with the current state of the science. *Clinical Child & Family Psychology Review, 8*, 327–342.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Burman, M. E., & McKay, S. (2007). Marginalization of girl mothers during reintegration from armed groups in Sierra Leone. *International Nursing Review, 54*, 316–323.

Canino, G., & Alegria, M. (2008). Psychiatric diagnosis—Is it universal or relative to culture? *Journal of Child Psychology and Psychiatry, 49*, 237–250.

Coalition to Stop the Use of Child Soldiers. (2004). *Child soldiers: Global report 2004*. London: Author.

Coalition to Stop the Use of Child Soldiers. (2008). *Child soldiers: Global report 2008*. London: Author.

Davison, E. H., Pless, A. P., Gugliucci, M. R., King, L. A., King, D. W., Salgado, D. M., et al. (2006). Late-life emergence of early-life trauma. *Research on Aging, 28*, 84–114.

Derluyn, I., Broekaert, E., Schuyten, G., & De Temmerman, E. (2004). Post-traumatic stress in former Ugandan child soldiers. *Lancet, 363*(9412), 861–863.

Dohrenwend, B. S., & Dohrenwend, B. P. (Eds.). (1974). *Stressful life events: Their nature and effects*. New York: Wiley.

Elbedour, S., Bensel, R., & Bastien, D. T. (1993). Ecological integrated model of children of war: Individual and social psychology. *Child Abuse and Neglect, 17*, 805–819.

Eyber, C., & Ager, A. (2003). Poverty and displacement: Youth agency in Angola. In S. Carr & T. Sloan (Eds.), *Community psychology and global poverty*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Grossman, D. (1996). *On killing: The psychological cost of learning to kill in war and society*. New York: Back Bay Books.

Guterman, N. B., & Cameron, M. (1997). Assessing the impact of community violence on children and youths. *Social Work, 42*, 495–505.

Guterman, N. B., Cameron, M., & Staller, K. (2000). Definitional and measurement issues in the study of community violence among children and youths. *Journal of Community Psychology, 28*, 571–587.

Hatch, S. L., & Dohrenwend, B. P. (2007). Distribution of traumatic and other stressful life events by race/ethnicity, gender, SES and age: A review of the research. *American Journal of Community Psychology, 40*, 313–332.

Hofmann, S. G., Litz, B. T., & Weathers, F. W. (2003). Social anxiety, depression, and PTSD in Vietnam veterans. *Journal of Anxiety Disorders, 17*, 573.

Honwana, A. (2006). *Child soldiers in Africa*. Philadelphia: University of Pennsylvania Press.

Hundt, G. L., Chatty, D., Thabet, A. A., & Abuateya, H. (2004). Advocating multi-disciplinarity in studying complex emergencies: The limitations of a psychological approach to understanding how young people cope with prolonged conflict in Gaza. *Journal of Biosocial Science, 36*, 417.

Johnson, K., Asher, J., Rosborough, S., Raja, A., Panjabi, R., Beadling, C., et al. (2008). Association of combatant status and sexual violence with health and mental health outcomes in postconflict Liberia. *JAMA, 300*, 676–690.

King, D. W., & King, L. A. (1999). Posttraumatic stress disorder in a national sample of female and male Vietnam veterans: risk. *Journal of Abnormal Psychology, 108*, 164.

King, L. A., King, D. W., Fairbank, J. A., Keane, T. M., & Adams, G. A. (1998). Resilience-recovery factors in post-traumatic stress disorder among female and male Vietnam veterans: Hardiness, postwar social support, and additional stressful life events. *Journal of Personality and Social Psychology, 74*, 420–434.

Kohrt, B. (2007). *Recommendations to promote psychosocial well-being of children associated with armed forces and armed groups (CAAFAG) in Nepal*. Kathmandu: UNICEF.

Kohrt, B. A., Jordans, M. J. D., Tol, W. A., Speckman, R. A., Maharjan, S. M., Worthman, C. M., et al. (2008). Comparison of mental health between former child soldiers and children never conscripted by armed groups in Nepal. *JAMA, 300*, 691–702.

Layne, C. M., Warren, J. S., Watson, P. J., & Shalev, A. Y. (2007). Risk, vulnerability, resistance, and resilience. In M. J. Friedman (Ed.), *Handbook of PTSD* (pp. 497–520). New York: Guilford.

Lazarus, R. S., & Folkman, S. (1984). *Psychological stress and the coping process*. New York: Springer.

Loughry, M., Ager, A., Flouri, E., Khamis, V., Hamid, A. H., & Qouta, S. (2006). The impact of structured activities among Palestinian children in a time of conflict. *Journal of Child Psychology and Psychiatry, 47*, 1211–1218.

Lustig, S. L., Kia-Keating, M., Knight, W. G., Geltman, P., Ellis, H., Kinzie, J. D., et al. (2004). Review of child and adolescent refugee mental health. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 24–36.

Luthar, S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543–562.

Luthar, S. S., & Goldstein, A. (2004). Children's exposure to community violence: Implications for understanding risk and resilience. *Journal of Clinical Child & Adolescent Psychology, 33*, 499–505.

Machel, G. (1996). *Impact of armed conflict on children*. New York: United Nations.

Machel, G. (2001). *The impact of war on children*. London: Hurst & Company.

Mac-Ikemenjima, D. (2008). Youth development, reintegration, reconciliation and rehabilitation in post-conflict West Africa: A framework for Sierra Leone, Liberia and Cote d'Ivoire. *International NGO Journal*, 3(9), 146–151.

Macksoud, M. S., & Aber, J. L. (1996). The war experiences and psychosocial development of children in Lebanon. *Child Development, 67*, 70–88.

MacMullin, C., & Loughry, M. (2004). Investigating psychosocial adjustment of former child soldiers in Sierra Leone and Uganda. *Journal of Refugee Studies, 17*, 460–472.

Maguen, S., Suvak, M., Litz, B. T., Adler, A. B., Castro, C. A., & Britt, T. W. (2006). *Predictors and prevalence of posttraumatic stress disorder among military veterans*. Westport, CT: Praeger Security International.

Masten, A. S., & Obradovic, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society*, 13, 1–16.

Matias-Carrelo, L. E., Chavez, L. M., Negron, G., Canino, G., Aguilar-Gaxiola, S., & Hoppe, S. (2003). The Spanish translation and cultural adaptation of five mental health outcome measures. *Culture, Medicine and Psychiatry, 27*, 291–313.

Mazurana, D., & Carlson, K. (2004). *From combat to community: Women and girls of Sierra Leone*. Washington, DC: Hunt Alternatives Fund, Initiative for Inclusive Security.

Mazurana, D., & Carlson, K. (2006). *The girl child and armed conflict: Recognizing and addressing grave violations of girls' human rights*. Florence, IT: United Nations.

Mazurana, D., & McKay, S. (2003). *Girls in fighting forces in Northern Uganda, Sierra Leone, and Mozambique: Policy and program recommendations*. Gatineau: Canadian International Development Agency.

McEwen, B. S. (2000). Allostasis and allostatic load: Implications for neuropsychopharmacology. *Neuropsychopharmacology, 22*, 108–124.

McKay, S., & Mazurana, D. (2004). *Where are the girls? Girls in fighting forces in Northern Uganda, Sierra Leone, and Mozambique: Their lives during and after war*.

Montreal, Canada: International Center for Human Rights and Democratic Development.

McKay, S., Robinson, M., Gonsalves, M., & Worthen, M. (2006). *Girls formerly associated with fighting forces and their children: Returned and neglected*. London: Coalition to Stop the Use of Child Soldiers.

National Committee for Demobilisation Disarmament and Reintegration. (2004). *Children and the armed conflict in Sierra Leone* (Chapter 4). Final Report of the Truth and Reconciliation Commission of Sierra Leone, Vol. 3b Retrieved from http://www.sierra-leone.org/TRC Documents.html

National Scientific Council on the Developing Child. (2005). *Excessive stress disrupts the architecture of the developing brain* (Working Paper 3). Cambridge, MA: The Center on the Developing Child, Harvard University.

Nelson, C. A., III, Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science, 318*(5858), 1937–1940.

Netland, M. (2005). Event-list construction and treatment of exposure data in research on political violence. *Journal of Traumatic Stress, 18*, 507–517.

Peters, K., & Richards, P. (1998). Why we fight: Voices of youth compatants in Sierra Leone. *Africa, 68*, 183–215.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*, 85–95.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rutter, M. (2006). Implications of resilience concepts for scientific understanding. In B. Lester, A. Masten & B. McEwen (Eds.), *Resilence in children. Annals of the New York Academy of Sciences* (Vol. 1094). Malden, MA: Blackwell.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin, 127*, 291–324.

Santacruz, M. L., & Arana, R. E. (2002). Experiences and psychosocial impact of the El Salvador civil war on child soldiers. *Biomedica, 22*(Suppl. 2), 283–397.

SAS Institute. (2004). *SAS* (Version 9.1.2). Cary, NC: Author.

Shepler, S. (2005). The rites of the child: Global discourses of youth and reintegrating child soldiers in Sierra Leone. *Journal of Human Rights, 4*, 197–211.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Smyke, A. T., Koga, S. F., Johnson, D. E., Fox, N. A., Marshall, P. J., Nelson, C. A., et al. (2007). The caregiving context in institution-reared and family-reared infants and toddlers in Romania. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 48*, 210–218.

Sommers, M. (2003). *Education in emergencies*. Washington, DC: Creative Associates International.

Tajima, E. A., Herrenkohi, T. I., Huang, B., & Whitney, S. D. (2004). Measuring child maltreatment: A comparison of prospective parent reports and retrospective adolescent reports. *American Journal of Orthopsychiatry, 74*, 424–435.

Thoits, P. A. (1991). On merging identity theory and stress research. *Social Psychology Quarterly, 54*, 101–112.

United Nations International Children's Emergency Fund. (2003). *Guide to the optional protocol on the involvement of children in armed conflict*. New York: Author.

United Nations International Children's Emergency Fund. (2005). *The disarmament, demobilization and reintegration of children associated with the fighting forces: Lessons learned in Sierra Leone*. New York: Author.

United Nations International Children's Emergency Fund. (2007). *Paris principles: Principles and guidelines on children associated with armed forces or armed conflict*. New York: Author.

Weiss, D., & Marmar, C. (1997). The impact of event scale-revised. In T. Keane (Ed.), *Assessing psychological trauma and PTSD* (pp. 399–411). New York: Guilford.

Wessells, M. (2005). Child soldiers, peace education, and postconflict reconstruction for peace. *Theory Into practice, 44*, 363–369.

Wessells, M. G. (2006). Child soldiering: Entry, reintegration, and breaking cycles of violence. In M. Fitzduff & C. Stout (Eds.), *The psychology of resolving global conflicts: From war to peace* (Vol. 3, pp. 243–266). Westport, CT: Praeger Security International.

Williamson, J. (2005). *Reintegration of child soldiers in Sierra Leone. January 31–February 9*. Washington, DC: U.S. Agency for International Development.

The World Revolution. (2001). *Child soldiers: A global problem*. Global Report on Child Soldiers 2001. Retrieved from http://www.worldevolution.org/projects/web guide/GuideArticle.asp?ID=9

Worthman, C. M., & Panter-Brick, C. (2008). Homeless street children in Nepal: Use of allostatic load to assess the burden of childhood adversity. *Development and Psychopathology, 20*, 233–255.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1096–1113

Posttraumatic Resilience in Former Ugandan Child Soldiers

Fionna Klasen
University Medical Center of Hamburg and
University of Hamburg

Gabriele Oettingen
New York University and
University of Hamburg

Judith Daniels
University Medical Center of Hamburg

Manuela Post and Catrin Hoyer
University of Hamburg

Hubertus Adam
University Medical Center of Hamburg

The present research examines posttraumatic resilience in extremely exposed children and adolescents based on interviews with 330 former Ugandan child soldiers (age = 11–17, female = 48.5%). Despite severe trauma exposure, 27.6% showed posttraumatic resilience as indicated by the absence of posttraumatic stress disorder, depression, and clinically significant behavioral and emotional problems. Among these former child soldiers, posttraumatic resilience was associated with lower exposure to domestic violence, lower guilt cognitions, less motivation to seek revenge, better socioeconomic situation in the family, and more perceived spiritual support. Among the youth with significant psychopathology, many of them had symptoms extending beyond the criteria for posttraumatic stress disorder, in keeping with the emerging concept of developmental trauma disorder. Implications for future research, intervention, and policy are discussed.

> *I want to get married, buy a bicycle, and put up a building.* (13-year-old boy, former child soldier)
>
> *I want to get a sewing machine so that I keep on making clothes and sell them to get some money.* (15-year-old girl, former child soldier)
>
> *I will be a person who is responsible in the community; I will be an honest person; I will be a person who helps people.* (16-year-old boy, former child soldier)

Note. Quotations stem from the interviews of the present study.

Millions of children around the globe suffer the consequences of armed conflicts. One of the most complex and severely traumatized group of war-affected children are child soldiers. Despite international bans, more than 250,000 children and adolescents are exploited as soldiers worldwide, almost half of them in Africa (Office of the Special Representative of the Secretary-General for Children and Armed Conflict, 2006). The term *child soldier* refers to any person under 18 years of age associated with an armed force or armed group in any capacity ranging from combatants to cooks (United Nations Children's Fund [UNICEF], 2007). During their time as soldiers, these youth are brutalized and cruelly abused by armed groups, and often forced to commit atrocities themselves. The literature on the psychological consequences of child soldiers' experiences is only now emerging and consists mainly of interview-based, ethnographic case studies (Betancourt et al., 2008). The ideological commitment of child soldiers (Kanagaratnam, Raundalen, & Asbjornsen, 2005), the situation of girl soldiers (McKay, Robinson, Gonsalves, & Worthen, 2006), the reintegration and psychological rehabilitation (Betancourt et al., 2008; Medeiros, 2007), and mental health status of former child

---

First and foremost, our gratitude goes to the children at Laroo Boarding School who made this research possible. We extend our gratitude to Malisa Mukanga and Rahel Duresso for their assistance with data collection, to Christophe Bayer for his creative ideas, and to Claus Barkmann, Monica Blotevogel, Tom Toepfer, Elizabeth Stephens, Bill Charette, Victoria Olsen, Monika Bullinger, and Georg Romer for their helpful comments on earlier drafts. This study was funded by the Children for Tomorrow Foundation.

Correspondence concerning this article should be addressed to Fionna Klasen, Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Center Hamburg, Martinistrasse 52, 20246 Hamburg, Germany. Electronic mail may be sent to f.klasen@uke.uni-hamburg.de.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0006

soldiers (Bayer, Klasen, & Adam, 2007; Derluyn, Broekaert, Schuyten, & De Temmerman, 2004; Kohrt et al., 2008) have been investigated.

One of the most impressive phenomena of child development is the ability of many children to develop into healthy, well-adapted adults despite adversity and trauma. The current research addresses this underresearched phenomenon of positive adaptation following exposure to extremely adverse conditions of armed conflict and domestic violence.

Traumatic Experiences in Children and Adolescents

Many experiences of children in armed conflict can be classified as traumatic. A stressor is defined as traumatic when it threatens an individual's life or physical integrity and elicits a subjective response of fear and helplessness (American Psychiatric Association, 2000). This definition includes both objective characteristics of the event and subjective responses on the victim's side. Terr (1991) proposed a framework for childhood trauma that distinguishes between single-incident trauma (Type I) and repeated or prolonged trauma (Type II). These Type II traumas have been also termed *complex trauma* (Herman, 1992; van der Kolk, 2005) and describe the ''experience of multiple, chronic and prolonged, developmentally adverse traumatic events, most often of an interpersonal nature (e.g., sexual or physical abuse, torture) and early-life onset'' (van der Kolk, 2005), which are characteristic of the experiences of child soldiers.

Such experiences have detrimental effects on children's development. When exposed to unpredictable and uncontrollable danger, a child will immediately react to frightening stimulus with fight, flight, or freeze response without being able to learn from the experience. Resources normally dedicated to growth and development are allocated to survival instead (van der Kolk, 2007). This means they may not be able to complete developmental tasks, for example, the development of secure attachment relationships, of a stable and integrated self-concept, and of the competence to self-regulate emotion and behavior (van der Kolk, 2005).

Multiple Posttraumatic Outcomes in Children and Adolescents

Psychopathological Outcome

Children exposed to traumatic events are at risk of developing psychopathological symptoms (Pynoos, Steinberg, & Wraith, 1995; Yule, 2002). Terr (1991) described four characteristics related to childhood trauma that appear to last for long periods of time, regardless of the diagnosis. These are: repeatedly intrusive memories of the traumatic event, repetitive behaviors, trauma-specific fears, and altered attitudes toward people, life, and the future. The introduction of a new psychiatric disorder termed *posttraumatic stress disorder* (PTSD) in 1980 was an attempt to establish a diagnostic category to describe trauma reactions in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Edition (*DSM–III*; American Psychiatric Association, 1980). The symptoms include repeated and unwanted reexperience of the event, hyperarousal, and avoidance of stimuli that could remind of the event. PTSD rapidly became the most frequently used outcome measure in the trauma literature.

Regarding prevalences of PTSD in former Ugandan child soldiers, two recent studies found that about one third (27%–34.9%) of Ugandan child soldiers developed the disorder according to *DSM–IV* criteria (Bayer et al., 2007; Okello, Onen, & Musisi, 2007). Other studies measured posttraumatic stress symptoms (PTSS) and indicated that almost all former Ugandan child soldiers (97%–98%) suffer from clinically significant PTSS (Amone-P'Olak, Garnefski, & Kraaij, 2007; Derluyn et al., 2004). These numbers seem much higher than the above reported prevalence rates on PTSD as they also include child soldiers with only moderate levels of symptoms. Even if not as thoroughly investigated as PTSD, other psychiatric sequelae are also common in war-affected children, particularly depression and anxiety disorders, dissociation, somatic complaints, as well as behavioral problems like aggressive and disruptive behavior (Qouta, Punamaki, Miller, & El-Sarraj, 2008; Shaw, 2003; Thabet, Abed, & Vostanis, 2004). Some studies on child soldiers concentrated solely on PTSD symptoms (Bayer et al., 2007; Derluyn et al., 2004), but recent research also indicates high prevalence rates for depression and anxiety (Kohrt et al., 2008; Okello et al., 2007).

When investigating war-traumatized children, two debates in the trauma literature are of crucial relevance. First, the diagnosis of PTSD might not integrate the full range of psychological sequelae following prolonged and repeated complex traumatic exposure (Herman, 1992), and second, the diagnosis of PTSD might not be developmentally sensitive (van der Kolk, 2005). Referring to both arguments, van der Kolk (2005) proposed the concept of developmental trauma disorder (DTD). This concept is close to what has been also described as *complex*

PTSD (Herman, 1992) or disorders of extreme stress not otherwise specified (DESNOS; Pelcovitz et al., 1997). Key features of DTD are emotional dysregulation, disturbed attachment patterns, behavioral reenactment (e.g., aggressive behavior), and persistently altered attributions and expectancies. Diagnostic criteria for DTD have been proposed by the Complex Trauma Task Force of the National Child Traumatic Stress Network (van der Kolk, 2007), but valid self-report measurements have yet to be developed. For assessing DTD, the Complex Trauma Task Force recommends a combination of PTSD measures and other measures of childhood psychopathology (e.g., Youth Self-Report [YSR]; Achenbach, Cook, Blaustein, Spinazzola, & van der Kolk, 2003). DTD has been proposed as new childhood trauma diagnosis for *DSM–V* (van der Kolk, 2005).

Resilient Outcome

Even in the face of severe adversity, some individuals show no signs of psychopathology. Already during the 1970s, the seminal work of Norman Garmezy, Michael Rutter, and Emmy Werner pointed out that a considerable proportion of children showed no pathology and high levels of competence, despite suffering risk experiences that would be expected to produce serious sequelae (Garmezy, Masten, & Tellegen, 1984; Rutter, 1987; Werner & Smith, 1982). This phenomenon is referred to as *resilience.*

There is an ongoing debate on whether resilience should be conceptualized as a personality trait or a mental health outcome (Luthar, Cicchetti, & Becker, 2000; Masten, 2007). Even when resilience is conceptualized in terms of outcome, conceptualizations vary in defining resilience as the absence of negative outcome, presence of positive outcome, or a combination of both. Recent approaches go even further and conceptualize resilience as a dynamic process when analyzing how risk and protective factors corroborate in modulating the impact of traumatic experiences on mental health outcome (Olsson, Bond, Burns, Vella-Brodrick, & Sawyer, 2003; Rutter, 1987).

In the face of highly adverse risk factors such as traumatic events, the absence of pathology is remarkable and may serve as the definition of resilience (Hoge, Austin, & Pollack, 2007; Vanderbilt-Adriance & Shaw, 2008). To date, only a few studies have been published examining resilience in trauma populations (Bonanno & Mancini, 2008). Instead, many studies on resilience investigated normative birth cohorts (Werner & Smith, 2001) or American middle-class samples (Criss, Pettit, Bates, Dodge, & Lapp, 2002). To understand resilience, however, it is also essential to focus on severely exposed samples, as resilience is not defined as a positive adjustment per se, but rather a positive adjustment in the context of high levels of adversity (Masten, 2001).

Due to heterogeneity in the operational definitions of adversity, adaptation, and resilience, the prevalence of resilience in high-risk samples is difficult to estimate. Vanderbilt-Adriance and Shaw (2008) concluded in their review that resilient outcomes above 25% are rare in higher risk samples (e.g., multiple risks, low socioeconomic status [SES]). In contrast, Bonanno, Galea, Bucciarelli, and Vlahov (2006) showed in a recent study that the prevalence of resilience in trauma survivors varied from 33% to 54% across different exposure categories.

Relating to these investigations, we emphasize the use of a relatively new term, *posttraumatic resilience*, to describe trauma survivors with a positive posttraumatic mental health outcome. A search of the PsychINFO database in June 2008 yielded 14,969 items referring to posttraumatic stress compared to only two items referring to posttraumatic resilience. In the absence of a definitive resilience research paradigm, Yehuda and Flory (2007) argued that investigators must clearly outline their definition of resilience. For the current research we defined posttraumatic resilience as the absence of clinically significant psychopathology in the aftermath of severe trauma exposure.

Risk and Protective Factors for Posttraumatic Mental Health Outcome

Masten, Cutuli, Herbers, and Reed (2009) distinguish between risk factors, promotive factors, and protective factors. A *risk factor* is defined as a measurable characteristic of a group of individuals or their situation that predicts negative outcome on a specific outcome criterion. *Promotive factors* and *protective factors* are defined as measurable characteristics of a group of individuals or their situation that predict positive outcome with respect to a specific criterion. While the first applies to factors that predict positive outcome in similar ways across risk levels, the latter refers to factors that predict positive outcome specifically in the context of risk or adversity. Although the predictors analyzed in the present research may predict better outcome in general, they appear to be particularly important under conditions of higher risk or adversity. We therefore refer to them as protective factors. Factors may relate to diverse domains (e.g., biological, psychological, sociological) and different levels within a domain (e.g., cognitive processes, personality traits; Kazdin, Kraemer, Kessler, Kupfer, & Offord, 1997).

In addition to trauma severity, female gender, low SES, and previous trauma or other childhood adversity are risk factors for posttraumatic psychopathology in children and adults (Brewin, Andrews, & Valentine, 2000; Pfefferbaum, 1997). In the absence of a posttraumatic resilience paradigm, variables can be drawn from two sets of research findings in the search for additional predictors. The resilience literature yields dispositional and social protective factors predicting positive outcomes and the trauma literature adds responses to trauma that are risk factors, as well as social support variables as protective factors for psychopathological outcomes.

Dispositional factors. Resilience research has a long-standing tradition of investigating dispositional attributes of well-adjusted children in the face of adversity (see Masten et al., 2009, for a review). Among other attributes, internal locus of control, a sense of meaning, social problem-solving skills, and strong self-esteem often have been implicated as protective factors for resilience (Masten et al., 2009; Rutter, 1985; Taylor, Kemeny, Reed, Bower, & Gruenewald, 2000). The concept of *hardiness* integrates many of these characteristics. It describes individuals who are committed to finding meaningful purpose in life, believe that one can influence one's surroundings and the outcome of events, and trust that one can learn and grow from both positive and negative life experiences (Kobasa, Maddi, & Kahn, 1982). Hardiness is associated with posttraumatic adjustment (King, King, Fairbank, Keane, & Adams, 1998). Furthermore, a positive orientation toward the future, comprising optimism and goal orientation, also is associated with good psychological adjustment in the context of adversity (Masten et al., 2009). Optimism is defined as the global expectation that the future will bring good things (Peterson, 2000). These globally positive expectancies are considered a major determinant of whether people continue to pursue valued life goals against the backdrop of adversity.

Parents and relationships. The resilience literature also strongly implicates the protective influence of parents and other supportive relationships for resilience in children and youth (Luthar, 2006; Masten et al., 2009). In the context of exposure to severe adversity, children in close proximity to effective caregivers fare better and young people who lose their parents are more vulnerable to psychopathology. Positive relationships with parents and other adults appear to provide emotional security as well as many forms of direct assistance during severe adversity. As children grown older, relationships with friends and romantic partners increase in importance. One of the most devastating effects of adversity is the destruction or loss of the fundamental protections afforded by caring adults and friends, such as when parents are killed.

Responses to trauma and perceived social support factors. Cognitive trauma theorists have focused on cognitive risk factors to explain the development and maintenance of posttraumatic psychopathology in adults and children (Ehlers & Clark, 2000; Ehlers, Mayou, & Bryant, 2003). The experience of traumatic events often changes the victim's thoughts and beliefs. For example, trauma survivors frequently experience guilt and self-blame (Kubany, 1994; Kubany et al., 1996). This guilt cognition can be related to having survived while others did not. It may also arise from a lack of justification of certain behaviors enacted during traumatic situations (Kubany, 1994). Guilt has been identified in trauma survivors such as victims of rape, battered women, and combat veterans (Kubany, Abueg, Kilauano, Manke, & Kaplan, 1997). Trauma survivors may also become preoccupied with feeling of unfairness and injustice, which may provoke rumination focusing on revenge (Orth, Montada, & Maercker, 2006). These responses can act as a block to cognitive change and thus serve to maintain PTSD. Further, aspects of the peritraumatic psychological response, predominantly dissociations, have been shown to be risk factors for posttraumatic outcome in adults and children (Brewin et al., 2000; Cook et al., 2003; Diseth, 2005; Ozer, Best, Lipsey, & Weiss, 2003). Finally, *perceived social support* has been described as a crucial protective factor for posttraumatic outcome (Bonanno, Galea, Bucciarelli, & Vlahov, 2007; Ozer et al., 2003). Perceived social support refers to the belief that specific transactions such as helping behavior would be provided by others if needed (e.g., by parents, teachers, friends). In the aftermath of trauma, perceived social support seems to serve an important secondary prevention role (Ozer et al., 2003).

The Present Research

Since the late 1980s Northern Uganda has been ravaged by the rebel terror of the Lord's Resistance Army (LRA). Eighty percent (1.4 million) of the population lives in camps for internally displaced people, many have been killed or tortured, and an estimated 25,000 children and adolescents have been forcefully recruited into the rebel forces (Coalition to Stop the Use of Child Soldiers, 2008). An estimated 24% of them are girls (Pham, Vinck, & Stover, 2007). Peace talks between the Ugandan government and the LRA began in July 2006 in southern Sudan. In November 2006, when this research was carried out, the atmosphere was tense with both hope for peace and fear of another failure in attaining a truce. While the total number of remaining LRA fighters is unknown, up to 2,000 women and children are believed to still remain with the rebel fighters (Coalition to Stop the Use of Child Soldiers, 2008).

In this study, we address several gaps in the literature on child soldiers. There is very little empirical research on adaptation and resilience in child soldiers, particularly with a focus on multiple risk and protective factors. We distilled four clusters of risk and protective factors both from resilience and trauma literature that may explain posttraumatic outcome over and above trauma severity during abduction. First, additional trauma and life stressors have been shown to be significant risk factors for posttraumatic outcome (Brewin et al., 2000). We therefore assessed the most critical additional trauma and life stressors for our sample, namely the loss of parents and exposure to domestic and community violence. Second, we addressed dispositional factors, such as hardiness and optimism, that are presumed to promote posttraumatic resilience (Kobasa et al., 1982; Scheier & Carver, 1992). Third, according to the trauma literature and previous research on child soldiers, dissociation, guilt, and revenge are presumed to have an impact on posttraumatic outcome (Bayer et al., 2007; Diseth, 2005). Finally, in addition to the assessment of social support, a known protective factor in trauma populations (Ozer et al., 2003), we assessed perceived spiritual support during crisis. As resilient individuals may experience some form of transient stress reaction following trauma exposure, we investigated former child soldiers living in the safe environment of a boarding school, protected from rebel attacks and domestic and community violence for 4 months prior to data collection.

Hence, we had four hypotheses for this study. First, we expected that a considerable proportion of the investigated child soldiers would be resilient according to our operational definition of resilience based on their mental health status: (a) not meeting criteria for PTSD, (b) not meeting criteria for depression, or (c) scoring below the multicultural cutoff for behavioral and emotional problems. Second, we hypothesized that trauma severity during abduction would be a significant risk factor for posttraumatic outcome. Third, we hypothesized that loss of parents and domestic and community violence significantly would predict posttraumatic outcome. Finally, we hypothesized that person variables (dispositional variables, responses to trauma, perceived external support) would predict posttraumatic outcome over and above what can be predicted from trauma severity.

Method

Participants

The sample of this cross-sectional field study consisted of 330 former Ugandan child soldiers (48.5% girls). Children were on average 14.44 years old (*SD* = 1.57, range = 11–17 years). The children's ethnicity was mainly Acholi (61.5%) and Langi (38.2%), and they originated from the five war-torn northern districts of Uganda (Apac 18.2%, Gulu 21.8%, Kitgum 17.3%, Lira 20.6%, Pader 20.6%, Other 1.2%). Luo was the first language of all children. Ninety-eight percent of the children were Christians. Their fathers' occupations were mainly subsistence farmer (63.6%), trader (17.1%), or security and military servant (12.0%). Children had on average five biological siblings (range = 0–12).

Procedures

We recruited participants from a boarding primary school in Gulu Town in Northern Uganda, a special-needs school designed to support war-traumatized children, established by the government of Uganda. The only one of its kind in the entire country, the school aims to help war-traumatized children reach the same curriculum level as students in the public school system. Due to financial problems, no specialized interventions, like psychosocial programs, could be started at the school prior to our data collection. The selection of children for admission at the school required the assessment of traumatic war experiences and related symptoms with a questionnaire carried out by trained assessors, giving priority to extremely war-affected children. Community leaders confirmed war involvement of children during the selection process. Children were picked in almost equal proportions from the five war-torn northern districts of Uganda. The selection process required that 60% of the school population be female. The school opened in July 2006, meaning children had been in the safe environment of the boarding school, protected from rebel attacks and domestic and community violence for 4 months prior to data collection.

Written informed consent from the local authorities and oral informed consent from students and teachers were obtained for the study. Children were eligible if they had been a child soldier for at least 1 month, had returned from the armed group at least 6 months prior to the study, and were currently between 11 and 17 years of age. Children's abduction status was confirmed by teachers. Inclusion criteria were met by 358 students of the school. Three children declined to participate and were not asked to justify their decision. Children were assured that they were free to withdraw from the study at any point. None withdrew during data collection. Likert scales were explained and rehearsed in detail with all children. According to teachers' assessment, 45.9% of the children were illiterate. Questionnaires were read aloud to illiterate children by trained interviewers, while trained interviewers helped literate children fill out the questionnaires in class. The need for explanation and examples transformed the self-report method into an interview. Psychiatrists from Gulu Hospital were available to provide psychological support at all points of data collection, although support was never necessary. The school received books for a library as honoraria for supporting the assessment, while interviewers received payment. Data from 25 pupils could not be analyzed due to missing inclusion criteria or invalid data, resulting in a final population of 330 pupils for analysis.

The study was approved by the ethics committee of the Medical Association of Hamburg and by the Uganda National Council for Science and Technology.

Measures

Lacking well-validated concepts and measures based on research with Ugandan populations, we adapted concepts and measures from the trauma and resilience literature. The adaptability of these measurements was discussed with experienced mental health professionals and anthropologists originating from Uganda or who had done research in the country. All instruments were translated and back-translated into the local language (Luo) by linguists at the Institute of Languages at Makarere University in Kampala. The translators and the study director collectively resolved problems in the translated version by comparing the original English version with the back-translation. The administration of all materials was tested in a pilot phase for acceptance, adequacy, and applicability.

Sociodemographic Variables

We assessed age, gender, and SES of the participant's families. SES was operationalized by parents' educational levels and families' material resources. Items were taken from the Child War Trauma Questionnaire (CWTQ; Macksoud, 1992). Educational level of each parent was assessed by one item scored 0 (*can't read or write*), 1 (*has had some schooling*), 2 (*has completed primary school*), 3 (*has completed secondary school or vocational school*), or 4 (*holds a college degree*). Family material resources were assessed by four items pertaining to four categories of resource (i.e., food, shelter, clothing, and income) rated on scale from 0 (*very poor*) to 4 (*very rich*). These six items were added to a sum score. Higher scores indicate higher SES. The internal consistency for the current sample was Cronbach's alpha = .78. These simple measures of SES and educational level were discussed with experts in the field and seemed to have high face validity for the current sample.

Trauma Severity During Abduction

By definition, children who were abducted to serve as child soldiers experienced a traumatic event. To quantify trauma severity during abduction, we measured the reported number of traumatic events. We define both the experience of being a victim as well as that of being a perpetrator of violence as traumatic events, as children were forced to commit atrocities. Our Child Soldiers Trauma Questionnaire (CSTQ) is a 19-item yes–no statement questionnaire based on our own previous research on Ugandan child soldiers (Bayer et al., 2007) and the CWTQ (Macksoud, 1992). The CSTQ consisted of two subscales: The 13-item Victim Subscale (i.e., abduction; exposure: shooting, bomb explosion, massacre, air raid; deprivation: food, water; witness: injury, killing; victim: death threats, beatings, injury, rape) and the six-item Perpetrator Subscale (i.e., perpetrator: fighting, looting, abduction, torture, injury, killing). The internal consistency for each subscale was high (Cronbach's alpha = .73 for the Victim Subscale and Cronbach's alpha = .76 for the Perpetrator Subscale).

Posttraumatic Mental Health Outcome

To assess *PTSD* and *depression* we used the two accordant modules of the MINI-KID. The MINI-KID is the child and adolescent version of the Mini International Neuropsychiatry Interview (MINI; Sheehan et al., 1998), a short structured interview with axis 1 diagnostic categories of the *DSM–IV*. The MINI showed good reliability and validity coefficients compared to the Composite International Diagnostic Interview and to the Structured Clinical Interview for *DSM–IV* (Sheehan et al., 1998) and has recently been used for the first time in a study in Uganda (Okello et al., 2007). In addition to PTSD and depression we screened for a broad variety of *behavioral and emotional problems* using the YSR (Achenbach & Rescorla, 2001). The YSR is a widely used self-report instrument containing 105 problem items (e.g., ''I worry a lot,'' ''I destroy things belonging to others''). The scale consists of the following syndrome clusters: anxious-depressed, withdrawn-depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior. Items are rated 0 (*not true*), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*) on the basis of the preceding months. All items summed up form a Total Problems score. The YSR has been shown to have excellent psychometric qualities (Achenbach & Rescorla, 2001). The YSR has been used in over 6,000 publications reporting findings from 65 countries (Achenbach & Rescorla, 2007). The multicultural application of the YSR is based on the vast amount of multicultural research and enables users to display problem scale scores in relation to sets of norms based on societies that were found to have relatively low scores (Group 1), medium scores (Group 2), or high scores (Group 3; Achenbach et al., 2008). Group 2 of the multicultural norms, which are also labeled *standard norms*, are recommended if norms for a certain population are missing (Rescorla et al., 2007). The only published African norms for the YSR are for Ethiopia, which are allocated to Group 2 (Mulatu, 1997). We used Group 2 norms based on the above arguments. Cronbach's alpha for the current sample was .95 for the Total Problems score.

Operationalization of Posttraumatic Resilience

As noted earlier, posttraumatic resilience was defined by the absence of psychopathology operationalized by (a) not meeting criteria for PTSD, (b) not meeting criteria for depression, and (c) scoring below the multicultural cutoff for behavioral and emotional problems (YSR).

Additional Trauma and Life Stressors

Loss of parents was assessed by asking whether parents were deceased. Thus, *loss of parents* was defined as losing mother, father, or both parents through death, no matter what the cause.

Exposure to *domestic and community violence* was assessed with a self-developed checklist based upon a recent report on violence against children in Uganda (Naker, 2005). This report refers to experiences and perspectives on violence against Ugandan children of 1,406 children and 1,093 adults based on a survey using complementary research methods (i.e., questionnaires, focus group discussions, narrative role plays, key informant interviews). Four main areas of violence against children were identified: physical, emotional, sexual, and economic. Our checklist contained 19 yes–no statements with 13 items pertaining to violence experienced at home (i.e., physical violence: caning, burning, locking up; emotional violence: shouting, insulting, threatening, ignoring; sexual violence: sexual insults, sexual touches, rape; economic violence: denying food, overworking, denying money for school fees and health care) and 6 items concerning violence experienced in the community including school (i.e., physical violence: caning, burning, locking up; sexual violence: sexual insults, sexual touches, rape). The internal consistency for the current sample was Cronbach's alpha = .82.

Dispositions

To assess *hardiness* we used the 10-item version of the Connor–Davidson Resilience Scale (CD–RISC; Connor & Davidson, 2003; Campbell-Sills & Stein, 2007). Items are rated on a 5-point Likert scale ranging from 0 (*not true at all*) to 4 (*true nearly all the time*), with higher scores reflecting greater abilities to tolerate experiences such as change, personal problems, illness, pressure, failure, and painful feelings (e.g., ''I am able to adapt when changes occur,'' ''I deal with whatever comes my way,'' ''I tend to bounce back after illness, injury, or other hardships''). The 10-item version displayed good psychometric properties and is highly correlated with the original 25-item version (r = .92; Campbell-Sills & Stein, 2007). Internal consistency for the current sample was Cronbach's alpha = .75.

To assess *positive future orientation* we used the Positive Future Orientation Subscale of the Adolescent Resilience Scale (ARS; Oshio, Kaneko, Nagamine, & Nakaya, 2003). It contains five items referring to optimism and goal orientation rated on a 5-point Likert scale (e.g., ''I feel positive about my future,'' ''I strive toward my future goals''). The subscale has been found reliable and valid in other studies (Oshio et al., 2003). Internal consistency for the current sample was alpha = .73.

Responses to Trauma

To assess *peritraumatic dissociation*, the Peritraumatic Dissociative Experiences Questionnaire was used (RAND PDEQ; Marshall, Orlando, Jaycox, Foy, & Belzberg, 2002). The eight-item instrument, rated on a scale ranging from 0 (*not at all true*) to 4 (*extremely true*), asks respondents to what extent they experienced depersonalization, derealization, and altered sensations at the time of the traumatic events (e.g., ''What was happening didn't seem real, like I was in a dream or watching a movie''). The RAND PDEQ has demonstrated good internal consistency and test–retest reliability, as well as discriminant and convergent construct validity in both community violence survivors and individuals who experienced sexual abuse (Marshall et al., 2002). The internal consistency for the current sample was Cronbach's alpha = .82.

To assess cognitive aspects of *guilt* we used the 22-item Guilt Cognitions Scale of the Trauma-related Guilt Inventory (TRGI; Kubany et al., 1996). Items are scored on a 5-point scale ranging from 0 (*never* or *not at all true*) to 4 (*always* or *extremely true*; e.g., ''I blame myself for what happened''). The TRGI has high internal consistency and adequate temporal stability (Kubany et al., 1996). TRGI scales were significantly correlated with other measures of guilt in validity studies with Vietnam veterans and battered women (Kubany et al., 1996). The internal consistency for the current sample was Cronbach's alpha = .78 for the Guilt Cognitions Scale.

Participants' motivation to seek *revenge* against transgressors was measured by the Revenge Motivations Subscale of the Transgression-Related Interpersonal Motivations Scale (TRIM; McCullough et al., 1998). The revenge subscale comprises five items that measure motivation to seek revenge (e.g., ''I'll make him or her pay back''). Higher scores indicate higher revenge. The subscale has a high internal consistency ($\alpha = .90$; McCullough et al., 1998) and evidence of construct validity (McCullough, Bellah, Kilpatrick, & Johnson, 2001; McCullough et al., 1998). The internal consistency for the current sample was Cronbach's alpha = .81.

Perceived External Support

Perceived social support during stressful situations was measured by the item ''I have at least one close and secure relationship which helps me when I am stressed.'' The item was rated on a 4-point scale, higher scores indicating higher support. We assessed *perceived spiritual support* by asking children to respond to the following item ''During stressful situations I wonder whether God has abandoned me.'' This item was taken from a measure of spiritual coping (RCOPE; Pargament, Smith, Koenig, & Perez, 1998). The item was rated on a 4-point scale and reverse coded. Hence, higher scores indicate higher perceived spiritual support.

Results

Descriptive Statistics

Experiences During Abduction

On average, children were abducted at 10.75 years of age (*SD* = 2.30, range = 5–16 years). All children in the sample were abducted by rebel attacks, one exception being a child born in captivity. Different tasks were assigned to children during their captivity: 41.8% were assigned primarily front-line tasks, for example, fighting, looting, abducting civilians (55% for boys and 26.8% for girls); 28% performed mainly logistic tasks, for example, carrying loads, spying, escorting commanders (34.3% for boys and 20.8% for girls); and 27.7% were assigned domestic chores, for example, cooking, caring for younger children (10.1% for boys and 47.7% for girls). The children experienced several traumatic events. For example, 90.6% of the children were beaten by armed forces, 87.9% witnessed murder, 86.4% were threatened with death, and 25.8% were raped by members of the armed group (22.4% for boys and 29.4% for girls). The average score on the Victim Subscale of CSTQ was 10.26 (*SD* = 2.30, range = 1–13). Many children were forced to commit atrocities. For example, 65.2% looted houses, 59.1% abducted other children, and 52.6% killed another person. The average score on the Perpetrator Subscale of CSTQ was 3.36 (*SD* = 2.00, range = 0–6).

The average time spent in abduction was 19.81 months (*SD* = 17.72, range = 1–108) and children had returned from the armed group 31.75 months prior to data collection (*SD* = 17.86, range = 6–105). Under the risk of capture and death, 81.2% of the children had escaped the rebel forces. More than one third (38.6%) of the children had been abducted more than once, in some cases up to five times.

Additional Trauma and Life Stressors

Before entering the boarding school, 97.2% of the children had lived with family members or relatives. Many children had lost one or both parent(s): 43.1% were double orphans and 36.7% were single orphans. Children reported frequent experiences of domestic and community violence (88.8%). On a scale of 19 experiences, children reported on average 5.43 violent acts (*SD* = 4.03, range = 0–19).

Posttraumatic Outcome

More than every fourth child (27.6%) showed a resilient mental health outcome. Thus, 72.4% showed significant symptoms of psychopathology at outcome, including 33% who met *DSM–IV* criteria for PTSD and 36.4% who met criteria for major depression (18.5% qualified for diagnoses of both disorders). Compared to multicultural norms, 61.2% of the children scored above the cutoff for behavioral and emotional problems. No significant gender differences regarding mental health outcomes were found, PTSD: $\chi^2(1, N = 330) = 0.04$, $p = .84$; depression: $\chi^2(1, N = 330) = 0.17$, $p = .73$; and behavioral and emotional problems: $t(328) = .11$, $p = .91$. We split the sample in two age groups: younger adolescents (11–13 years, $n = 92$) and older adolescents (14–17 years, $n = 234$). Younger adolescents displayed significantly lower rates of PTSD and depression than older adolescents, PTSD: 23.9% versus 36.8%, $\chi^2(1, N = 330) = 4.91$, $p = .027$; depression: 27.2% versus 39.3%, $\chi^2(1, N = 330) = 4.23$, $p = .04$.

Mean values and standard deviations of psychological predictors of posttraumatic outcome are presented in Table 1. Table 2 displays zero-order correlation among all key variables. To control for possible conceptual overlap between responses to trauma and posttraumatic outcome in this study, the guilt items were removed from the outcome scales. Correlations were virtually as strong as with the full scales, indicating that the correlation was not a methodological artifact.

Multivariate Analysis Predicting Posttraumatic Resilient Outcome

Risk and Protective Factors

Hierarchical multiple logistic regression analysis was used to test the association of risk and protective factors with posttraumatic status. The rationale for entry order was as follows: In Step 1, we tested the association of sociodemographic variables with posttraumatic outcome and controlled these variables for further analysis steps. In Steps 2 and 3, we tested the hypothesized association of trauma severity during abduction and additional trauma and life stressors with posttraumatic outcome. In Step 4, we tested whether person variables predict posttraumatic outcome over and above what can be predicted on the basis of sociodemographic variables and trauma severity.

Table 1
Mean Values and Standard Deviations of Trauma, Person, and Outcome Variables

	All (*N* = 330)		Clinical group (*n* = 239)		Resilient group (*n* = 91)		
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p* value[a]
Trauma severity—Victim [0–13]	10.26	2.30	10.52	2.14	9.57	2.57	.001
Trauma severity—Perpetrator [0–6]	3.36	2.00	3.55	1.99	2.87	1.95	.005
Domestic and community violence [0–19]	5.43	4.03	6.22	4.07	3.36	3.10	< .001
Hardiness [0–40]	22.66	8.28	23.12	7.81	21.46	9.33	.083
Positive future orientation [5–25]	21.29	3.52	21.03	3.68	22.00	2.99	.045
Peritraumatic dissociation [8–40]	24.21	8.04	25.33	7.91	21.26	7.69	< .001
Guilt cognitions [0–88]	39.51	14.50	41.47	13.27	34.34	16.31	< .001
Revenge motivation [5–25]	13.01	5.76	13.90	5.86	10.66	4.74	< .001
Perceived social support [0–4]	1.76	1.06	1.65	1.05	2.05	1.04	.001
Perceived spiritual support [0–4]	1.58	1.30	1.35	1.25	2.18	1.23	< .001
PTSD symptoms [0–16]	8.89	3.37	10.01	2.86	5.97	2.80	< .001
Depression symptoms [0–13]	6.54	3.26	7.69	2.86	3.54	2.13	< .001
Total Problems score (YSR) [0–210]	75.64	33.21	88.15	29.48	42.78	15.07	< .001

Note. PTSD = posttraumatic stress disorder; YSR = Youth Self-Report.
[a] Differences between clinical and resilient group, Mann–Whitney *U* test.

Table 2

Zero-Order Correlations Among Key Variables

	Measure																		
Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Female gender																			
2. Age	−.29***																		
3. Family SES	.06	−.02																	
4. Age at abduction	−.13*	.31***	.00																
5. Length of abduction	−.02	.19**	−.13*	−.12*															
6. Trauma severity—Victim	−.12*	.23***	−.09	−.05	.26***														
7. Trauma severity—Perpetrator	−.25***	.11	−.04	−.01	.25***	.53***													
8. Time since return	−.10	.17**	.09	−.07	−.12*	−.08	.01												
9. Loss of parents	−.02	.20***	−.12*	−.09	.15**	.14**	.15**	.05											
10. Domestic and community violence	−.08	.17**	−.14*	.03	.01	.24***	.19***	.11*	.16**										
11. Hardiness	−.05	.12*	−.09	.06	.08	.15**	.12*	−.03	.10	.00									
12. Positive future orientation	.12*	−.01	−.08	−.06	.00	.12*	−.05	−.17**	−.02	−.13*	.30***								
13. Peritraumatic dissociation	−.05	.05	−.08	.06	.01	.19**	.24***	.01	.06	.20***	.26***	.11*							
14. Guilt cognitions	−.09	.04	−.12*	−.04	−.02	.20***	.27***	.04	.10	.20***	.22***	.20***	35***						
15. Revenge motivation	−.07	.14*	−.15**	.05	.05	.11	.02	.08	.00	.15**	.12*	−.05	.17**	.16**					
16. Perceived social support	.04	−.08	.11*	−.04	.02	−.04	−.00	−.08	−.09	−.16**	.16**	.28***	−.02	−.04	−.19**				
17. Perceived spiritual support	.04	.00	.05	−.06	−.02	−.12*	−.18**	−.03	.02	−.07	−.10	−.02	−.22***	−.20***	−.17**	−.06			
18. PTSD symptoms	−.05	.23***	−.14*	.03	.06	.28***	.24***	.12*	.21***	.32***	.23***	−.01	.33***	.27***	.21***	−.10	−.19***		
19. Depression symptoms	−.06	.19**	−.25***	.05	.04	.32***	.22***	.12*	.15**	.33***	.17**	−.01	.27***	.22***	.32***	−.11*	−.27***	.63***	
20. Total Problems score (YSR)	−.01	.20***	−.07	−.01	.03	.25***	.21***	.17**	.13*	.39***	.12*	−.08	.28***	.30***	.22***	−.12*	−.19**	.55***	.54***

Note. $N = 330$. Zero-order correlations are represented by Spearman correlations for continuous predictor variables and point-biserial correlations for dichotomous predictor variables. PTSD = posttraumatic stress disorder; SES = socioeconomic status; YSR = Youth Self-Report.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

The regression model controlled length of time since returning from abduction.

Hence, in Step 1, gender, current age, family SES, age at abduction, and length of abduction were entered into the regression equation. These variables accounted for 10% of the variance, Nagelkerke $R^2 = .10$, $\chi^2(5, N = 330) = 13.94$, $p = .016$. In Step 2, we added trauma severity during abduction (i.e., number of traumatic events victim, traumatic events as perpetrator), which did not significantly improve the prediction of outcome, $R^2_{\text{change}} = .01$, $\chi^2_{\text{change}}(2, N = 330) = 2.36$, $p = .31$; $R^2 = .11$, $\chi^2(7, N = 330) = 16.30$, $p = .02$. In Step 3, we entered additional trauma and life stressors into the regression equation (i.e., loss of parents, domestic and community violence). The accuracy of prediction increased significantly to 24% variance explained, $R^2_{\text{change}} = .13$, $\chi^2_{\text{change}}(2, N = 330) = 20.37$, $p < .001$; $R^2 = .24$, $\chi^2(9, N = 330) = 36.67$, $p < .001$. In Step 4, we entered the person variables into the regression function: dispositional factors (i.e., hardiness, positive future orientation), responses to trauma (i.e., dissociation, guilt, revenge), and perceived external support (i.e., social, spiritual support). Step 4 resulted in a significant increase in explained variance to 43%, $R^2_{\text{change}} = .19$, $\chi^2_{\text{change}}(7, N = 330) = 34.18$, $p < .001$; $R^2 = .43$, $\chi^2(16, N = 330) = 70.86$, $p < .001$.

Most Critical Predictors

Searching for the most critical predictors, six variables of the regression model proved to be significant predictors of posttraumatic resilience. Controlling for all other variables, higher age was a risk factor (OR = 0.72) and high family SES was a protective factor (OR = 1.11). Domestic and community violence was a risk factor (OR = 0.81) as were strong guilt cognitions and revenge motivation (guilt: OR = 0.96; revenge: OR = 0.92). Finally, children's perceived spiritual support was a protective factor and almost doubled the odds of resilience as defined in this study (OR = 1.91). Regression coefficients of all predictors are displayed in Table 3.

Interaction Effects

A second hierarchical multiple logistic regression analysis was used to test for the interaction effect between gender and age, the interaction effect between trauma severity and age, as well as the interaction effects between trauma severity and each of the most important hypothesized protective factors (i.e., hardiness, positive future orientation, perceived social support, and perceived spiritual support) in the prediction of posttraumatic mental health outcome. To compute these analyses, we

Table 3
Hierarchical Multiple Logistic Regression Predicting Posttraumatic Resilient Outcome

Predictor	*B*	*SE*	Wald	*p* value	OR	95% CI for OR
Sociodemographic variables						
Gender	−.56	.46	1.49	.22	0.57	0.23–1.40
Age	**−.32**	**.15**	**4.41**	**.04**	**0.72**	**0.54–0.98**
Family SES	**.11**	**.05**	**4.15**	**.04**	**1.11**	**1.00–1.23**
Age at abduction	.06	.10	0.40	.53	1.07	0.88–1.29
Length of abduction	.01	.01	0.54	.46	1.01	0.98–1.04
Trauma severity during abduction						
Trauma severity—Victim	.10	.13	0.62	.43	1.11	0.86–1.42
Trauma severity—Perpetrator	−.07	.13	0.27	.60	0.93	0.72–1.21
Additional trauma and life stressors						
Loss of parents	.10	.52	0.04	.85	1.10	0.40–3.04
Domestic and community violence	**−.21**	**.06**	**12.17**	**.00**	**0.81**	**0.72–0.91**
Person variables						
Hardiness	.03	.03	0.85	.36	1.03	0.97–1.08
Positive future orientation	.04	.07	0.33	.57	1.04	0.91–1.19
Peritraumatic dissociation	−.01	.03	0.05	.82	0.99	0.94–1.05
Guilt cognitions	**−.04**	**.02**	**5.32**	**.02**	**0.96**	**0.93–0.99**
Revenge motivation	**−.08**	**.04**	**4.16**	**.04**	**0.92**	**0.85–1.00**
Perceived social support	.07	.22	0.10	.76	1.07	0.70–1.64
Perceived spiritual support	**.65**	**.17**	**14.22**	**.00**	**1.91**	**1.37–2.68**

Note. Model fit: $\chi^2(16, N = 330) = 70.86$, $p < .001$, Nagelkerke $R^2 = 43.2\%$. Controlled for time since return. Significant predictors ($p < .05$) indicated in boldface. SES = socioeconomic status.

composed an overall trauma severity measure of trauma during abduction, loss of parents, and domestic and community violence (Cronbach's alpha = .85). All continuous variables were centered prior to forming interaction terms. After controlling for main effects, we tested for interaction effects: first, the interaction of gender and age; second, of trauma severity and age; and third, of trauma severity and the four protective factors respectively, predicting posttraumatic outcome. None of the interaction terms was significantly associated with posttraumatic outcome.

Discussion

Child soldiers are one of the most severely traumatized populations of children and adolescents. Despite the extremely high level of adversity exposure, close to 30% of children showed a posttraumatic resilient outcome according to our operational definition. The prevalence rate of PTSD found in the presented study is consistent with other recent studies on child soldiers (Bayer et al., 2007; Okello et al., 2007). In addition to PTSD, results indicated high rates of depression and emotional and behavioral problems in these former child soldiers. These results, in line with growing evidence in the trauma literature, suggest that the concept of PTSD does not fully capture the reaction to repeated and prolonged interpersonal trauma of child soldiers and may be better described by concepts such as developmental trauma disorder (van der Kolk, 2005). Consideration of risk and protective factors added substantially to the prediction of posttraumatic status among these child soldiers, over and above sociodemographic variables and trauma severity during abduction. These findings show convergence and divergence with findings on resilience in Western populations.

Demographic and Trauma Variables

When controlling for all other variables, gender did not relate to resilience. This finding applied to all mental health outcome measures, that is, PTSD, depression, and behavioral and emotional problems. In contrast, most epidemiological studies in Western population samples indicate female gender as a risk factor for pathology in adolescents whether the samples were traumatized (Breslau & Anthony, 2007; Giaconia et al., 1995) or nontraumatized (Rescorla et al., 2007; Twenge & Nolen-Hoeksema, 2002). There is ongoing controversy on whether these gender differences are due to gender-based response tendencies, intrinsic vulnerability among females, or whether it might be explained by gender differences in type and extent of trauma exposure, especially referring to sexual trauma (Breslau, Chilcoat, Kessler, Peterson, & Lucia, 1999; Cortina & Kubiak, 2006).

Older children showed more mental health problems than younger children. This observation is in line with the literature on Western population samples, both traumatized and nontraumatized, also showing higher rates of symptoms in older than in younger children (Green, Korol, Grace, & Vary, 1991; Rescorla et al., 2007). Taken together, these findings might indicate a normative trend independent of trauma exposure. We did not find a significant Gender × Age interaction effect. This finding contrasts with findings of larger increases of symptoms with age for girls than boys in Western population samples (Rescorla et al., 2007; Twenge & Nolen-Hoeksema, 2002). Finally, in line with the literature on Western samples, children originating from families with higher SES were more likely to fall in the resilient group (Bonanno et al., 2007; Masten et al., 2009).

Based on the earlier literature on war-affected children, we expected trauma severity during abduction would be a crucial predictor of posttraumatic outcome (Mollica, McInnes, Poole, & Tor, 1998). However, both in our sample and in previous research on child soldiers, this was not the case (Bayer et al., 2007). Future research is required to investigate whether the assumed dose–response relation between trauma severity and posttraumatic outcome may not hold true for extremely exposed populations. This could be described as a threshold effect after which dose is so high that variation in dose no longer matters. In line with previous research on child soldiers, age at abduction and length of abduction were not significant predictors of posttraumatic outcome (Bayer et al., 2007; Derluyn et al., 2004).

Regarding additional trauma and life stressors, loss of parents was not a significant risk factor for resilience. Whether African extended family systems compensated for the loss of parents remains an open question. Domestic and community violence was identified as a significant risk factor for posttraumatic resilience. The return of children from the often extremely traumatizing experiences with the rebels to a violent and rejecting home environment may have led to an ultimate loss of trust in other people. There is some evidence that the exposure to war violence might brutalize societies and facilitate domestic abuse (Catani, Jacob, Schauer, Kohila, & Neuner, 2008). Therefore, violent behavior within families and communities might emerge as a long-term effect of warfare lasting beyond the end of armed conflict. Further research in domestic violence inflicted on children in postconflict countries is needed.

Person Variables

The frequent experiences of domestic and community violence may explain why perceived social support was not a significant predictor of posttraumatic resilience. In contrast, perceived spiritual support was a significant factor pertaining to resilience. Children who did not feel abandoned by God during crises showed significantly fewer symptoms. Spirituality and Christian religion are deeply rooted in social and cultural life of Northern Ugandans and may therefore be an important source of healing and reconciliation (Harlacher, Okot, Obonyo, Balthazard, & Atkinson, 2006). The LRA rebels target young children also because they are easily indoctrinated through redirecting their faith toward the rebels' aims (e.g., promising the children bulletproofing through spiritual rituals or forcing the children to commit atrocities while saying the name of God). The perception of God in Ugandan youth and the relevance of religion for coping with trauma need to be further investigated.

Dispositions (i.e., hardiness and positive future orientation) did not assist in explaining the variance in posttraumatic outcome. We employed a 10-item version of the CD–RISC (Campbell-Sills & Stein, 2007). Items such as "I can deal with whatever comes my way" are perhaps unsuitable for extremely exposed individuals such as child soldiers. It also is possible that the items of this scale may not reflect successful adaptation in children embedded in the Northern Ugandan culture.

Despite their severe trauma, children displayed high average scores of positive future orientation. This is noteworthy, as a negative future orientation (e.g., does not expect to have a career or a normal life span) has been described as an indicator of pathological trauma response (American Psychiatric Association, 2000). Although research is scarce, there are some indications of higher dispositional optimism in African populations (Eshun, 1999). Positive future orientation did not appear to play a protective role with respect to posttraumatic resilience.

Peritraumatic dissociation was not a significant risk factor for posttraumatic outcome. Understanding the function of dissociation in the process of trauma response in child soldiers requires further investigation. Dissociation is mostly used as a clinical concept and a predictor of subsequent PTSD (Diseth, 2005). Recent research, however, indicates that so-called emotional dissociation may protect extremely exposed individuals (Bonanno, 2004; Coifman, Bonanno, Ray, & Gross, 2007).

Our finding that revenge motivation was a risk factor for worse posttraumatic outcome is consistent with other research on child soldiers (Bayer et al., 2007). Moreover, when ruminating about someone who has caused them harm, people become more aggressive, which may translate into violent behavior (McCullough, Bono, & Root, 2007). Revenge motivation has been shown to be negatively associated with openness to reconciliation and forgiveness, also in child soldiers (Bayer et al., 2007; McCullough et al., 2001). Overcoming revenge motives may therefore be a crucial step toward sustainable peace building processes.

Further, children indicated that they frequently suffered from guilt, expressing that the time spent with the armed group caused them pain and suffering, and reported recurrent thoughts of their own acts being unjustified and unforgivable. In addition, survivor guilt may have played a role (e.g., guilt about surviving when others did not; American Psychiatric Association, 2000). Importantly, guilt was related both to being a victim of violence as well as to being a perpetrator (Table 2). Children had frequently been forced to torture or kill other individuals, sometimes even their relatives, to survive. The will to survive, obedience to orders, normalization of violence, and ideology may be the underlying psychological processes that enable children to become perpetrators (Wessells, 2006). Whether different psychological processes lead to guilt in victims versus in perpetrators is an interesting objective for future research. Responses to trauma, such as revenge and guilt, might sustain posttraumatic psychopathology in child soldiers as conceptualized in the cognitive model of the maintenance of PTSD by Ehlers and Clark (2000).

Limitations and Directions for Future Research

While the school presented a rare opportunity to assess a relatively large number of traumatized child soldiers in relation to potential risk and protective factors, the study had a number of limitations. We discuss six limitations, which may set directions for future research in this difficult environment.

First, we were not able to include positive behavioral outcome measures due to several methodological issues and logistic problems. For example, we had wanted to include academic achievement, but reports or remarks on students did not exist. Further, we were not able to obtain multiple perspectives on children's functioning, as parents were scattered throughout the northern part of Uganda and traveling was too dangerous. Teacher's perspectives could not be obtained as the classes were too big (60 students) to result in reliable data. Therefore we had to operationalize resilience, relying exclusively on the absence of symptoms. Clearly, our measures did not fully capture the individual differences in adaptation, motivation, and hope that we observed in these young people and that also are evident in the quotations we have included. Future studies should try to include positive indicators of adjustment (e.g., getting along with people, meeting expectations for chores or school, contributing to the community) and further investigate what defines resilience or an adaptive developmental outcome in ongoing- or postconflict contexts. Such a perspective needs to incorporate specific criteria for positive outcome for the various stages of child development (infancy and early childhood, middle childhood, and adolescence). In the developmental literature, positive outcome is often defined by successfully meeting age-related developmental tasks (e.g., early childhood: attachment to caregivers; middle childhood: school adjustment; adolescence: forming close friendships; Masten & Coatsworth, 1998). Suffering from posttraumatic psychopathology such as intrusive thoughts or depressive symptoms might hinder children in completing developmental tasks and therefore cause further psychological and social maladaptation.

Second, our study was cross-sectional, and we are not able to make any conclusions regarding stability and change in psychological well-being over time. Therefore, we could not distinguish between resilience, recovery, and delayed onset (Bonanno, 2004). Moreover, the onset of psychopathology may have not been posttrauma in some of the participating children. However, the duration and magnitude of symptoms within the different posttraumatic trajectories need further empirical investigation. Longitudinal research on child soldiers is urgently needed to answer these open developmental questions.

Third, for feasibility reasons, we used a convenience sample of former child soldiers. Having chosen a special-needs school for war-traumatized children might have biased the sample toward an overestimation of symptoms in child soldiers, as the admission process gave priority to extremely war-traumatized children. On the other hand, the special school may have boosted the prevalence of resilience in some of these young people. Additional research is needed with more representative samples and control groups.

Fourth, there were limitations in our assessment tools. We had to rely on Western-oriented questionnaires due to the lack of locally developed instruments. Another weakness was the single-item measurement of social and spiritual support, which presumably did not cover the constructs as a whole. It also was not feasible to include a full diagnostic assessment, and measures were limited to selected modules of the MINI–KID and the YSR. Thus, it is possible that some of the young people had symptoms and disorders that were not assessed by the measures used. Our research indicates the need for development and validation of cultural and context-sensitive measurements suitable for use in highly exposed populations. Assessment of cross-culture validity and reliability of standard psychiatric measurements are needed even without the usual "gold standards" (Bolton, 2001).

Fifth, our study focused on person variables. Certainly, familial (e.g., attachment, parenting) and social (e.g., friends) variables, as well as cultural factors (e.g., reconciliation ceremonies), may also be important influences on posttraumatic resilience. As recently suggested by Betancourt and Khan (2008), posttraumatic resilience needs to be understood from an ecological, developmental perspective that includes family, peers, schools, communities, and cultural and political belief systems. Within this approach, the roles of attachment relationships, caregiver health, resources and connection in the family, and social support available in peers and extended social networks need to be examined in war-affected children. Restoring damaged social ecology is fundamental to child development and psychological rehabilitation for war-affected children. Future research needs to integrate an ecological perspective.

Finally, it was not feasible to study resilience beyond the behavioral level. To understand the factors maintaining posttraumatic psychopathology, the neuroscience of posttraumatic stress may be important. Most research on the neurological underpinnings of PTSD identify the amygdala and hippocampus as key brain areas involved in the registration of potentially dangerous situations and in the later formation of memories of such events and cite the hypothalamic–pituitary–adrenal axis as a central agent in both the development of PTSD and its maintenance (Bremner, Elzinga, Schmahl, & Vermetten, 2008; Yehuda & LeDoux, 2007). Differentiating neurobiological correlates of PTSD and resilience following trauma exposure could contribute to explaining interindividual differences in trauma responses (Yehuda & Flory, 2007).

Implications for Interventions

Resilience research has a pragmatic mission, namely to provide principles of healthy development for successful intervention and treatment approaches (Masten & Coatsworth, 1998). Therefore, despite the scientific challenges posed by political instability in armed conflict situations, research is sorely needed to develop adequate measurements and tailored interventions. Our findings provide initial clues on factors to address in such programs. Domestic and community violence appeared to have a detrimental effect on the children's mental health over and above the traumatic events during abduction. Further, responses to trauma, such as guilt cognition, were negatively related to posttraumatic resilience. Moreover, it may be important to address revenge motives to break through the cycle of violence and work toward reconciliation.

While we focused on resilience, it is important to note that three in four child soldiers in our study were suffering significant levels of posttraumatic psychopathology. Interventional research in conflict areas is still scarce but very much needed. Cognitive behavioral treatment modules for the children (Onyut et al., 2005) as well as family-focused stress management programs may have beneficial effects on children and their family systems.

Our findings have implications for training and policy as well, because high levels of posttraumatic psychopathology in former child soldiers contrast with a psychiatrist-to-population ratio of 1:1.3 million in Uganda (Ovuga, Boardman, & Wasserman, 2007). Therefore, training of mental health professionals and a structural basis for sustainable intervention programs for children and adolescents in the region are urgently needed. Such programs should be based on knowledge on how to strengthen and sustain resilience in traumatized children.

Conclusions

Our findings indicate that despite severe trauma exposure, a critical proportion of former child soldiers showed a posttraumatic resilient outcome. Domestic and community violence, guilt cognitions, revenge motivation, and perceived spiritual support explained a substantial degree of variance in posttraumatic resilience over and above trauma severity during abduction. Knowledge of these factors provides initial clues on how to create interventions toward strengthening resilience in former child soldiers.

The present research benefited from two lines of research in the literature—one on resilience and one on trauma—that have not been fully integrated in research to date. Future research on severely traumatized children and adolescents would benefit from further integration of these perspectives, with the goal of an integrated paradigm of posttraumatic adaptation and resilience in ongoing- or postconflict contexts. Only through continued research on the processes leading to positive outcome can a full understanding of posttraumatic resilience be reached to inform intervention programs. The ongoing goal of research on posttraumatic resilience is to learn more effective strategies for protecting children and promoting their recovery from the developmental threats of trauma and violence.

References

Achenbach, T. M., Becker, A., Dopfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., et al. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry*, *49*, 251–275.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2007). *Multicultural supplement to the manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed.). Washington, DC: Author.

Amone-P'Olak, K., Garnefski, N., & Kraaij, V. (2007). Adolescents caught between fires: Cognitive emotion regulation in response to war experiences in Northern Uganda. *Journal of Adolescence*, *30*, 655–669.

Bayer, C. P., Klasen, F., & Adam, H. (2007). Association of trauma and PTSD symptoms with openness to reconciliation and feelings of revenge among former Ugandan and Congolese child soldiers. *Journal of the American Medical Association*, *298*, 555–559.

Betancourt, T. S., Borisova, I., Rubin-Smith, J., Gingerich, T., Williams, T., & Agnew-Blais, J. (2008). *Psychosocial adjustment and social reintegration of children associated with armed forces and armed groups: The state of the field and future directions.* Austin, TX: Psychology Beyond Borders.

Betancourt, T. S., & Khan, K. T. (2008). The mental health of children affected by armed conflict: Protective processes and pathways to resilience. *International Review of Psychiatry, 20,* 317–328.

Bolton, P. (2001). Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *Journal of Nervous and Mental Disease, 189,* 238–242.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist, 59,* 20–28.

Bonanno, G. A., Galea, S., Bucciarelli, A., & Vlahov, D. (2006). Psychological resilience after disaster: New York City in the aftermath of the September 11th terrorist attack. *Psychological Science, 17,* 181–186.

Bonanno, G. A., Galea, S., Bucciarelli, A., & Vlahov, D. (2007). What predicts psychological resilience after disaster? The role of demographics, resources, and life stress. *Journal of Consulting and Clinical Psychology, 75,* 671–682.

Bonanno, G. A., & Mancini, A. D. (2008). The human capacity to thrive in the face of potential trauma. *Pediatrics, 121,* 369–375.

Bremner, J. D., Elzinga, B., Schmahl, C., & Vermetten, E. (2008). Structural and functional plasticity of the human brain in posttraumatic stress disorder. *Progress in Brain Research, 167,* 171–186.

Breslau, N., & Anthony, J. C. (2007). Gender differences in the sensitivity to posttraumatic stress disorder: An epidemiological study of urban young adults. *Journal of Abnormal Psychology, 116,* 607–611.

Breslau, N., Chilcoat, H. D., Kessler, R. C., Peterson, E. L., & Lucia, V. C. (1999). Vulnerability to assaultive violence: Further specification of the sex difference in posttraumatic stress disorder. *Psychological Medicine, 29,* 813–821.

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology, 68,* 748–766.

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 20,* 1019–1028.

Catani, C., Jacob, N., Schauer, E., Kohila, M., & Neuner, F. (2008). Family violence, war, and natural disasters: A study of the effect of extreme stress on children's mental health in Sri Lanka. *BMC Psychiatry, 8,* 33.

Coalition to Stop the Use of Child Soldiers. (2008). *Child soldiers.* Global report 2008. Retrieved June 10, 2008, from http://www.child-soldiers.org/library/global-reports

Coifman, K. G., Bonanno, G. A., Ray, R. D., & Gross, J. J. (2007). Does repressive coping promote resilience? Affective-autonomic response discrepancy during bereavement. *Journal of Personality and Social Psychology, 92,* 745–758.

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor–Davidson Resilience scale (CD–RISC). *Depression and Anxiety, 18,* 76–82.

Cook, A., Blaustein, M., Spinazzola, J., & van der Kolk, B. (2003). *Complex trauma in children and adolescents* (White paper from the National Child Traumatic Stress Network Complex Trauma Task Force). Los Angeles: National Center for Child Traumatic Stress.

Cortina, L. M., & Kubiak, S. P. (2006). Gender and posttraumatic stress: Sexual violence as an explanation for women's increased risk. *Journal of Abnormal Psychology, 115,* 753–759.

Criss, M. M., Pettit, G. S., Bates, J. E., Dodge, K. A., & Lapp, A. L. (2002). Family adversity, positive peer relationships, and children's externalizing behavior: A longitudinal perspective on risk and resilience. *Child Development, 73,* 1220–1237.

Derluyn, I., Broekaert, E., Schuyten, G., & De Temmerman, E. (2004). Posttraumatic stress in former Ugandan child soldiers. *Lancet, 363,* 861–863.

Diseth, T. H. (2005). Dissociation in children and adolescents as reaction to trauma—An overview of conceptual issues and neurobiological factors. *Nordic Journal of Psychiatry, 59,* 79–91.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38,* 319–345.

Ehlers, A., Mayou, R. A., & Bryant, B. (2003). Cognitive predictors of posttraumatic stress disorder in children: Results of a prospective longitudinal study. *Behaviour Research and Therapy, 41,* 1–10.

Eshun, S. (1999). Cultural variations in hopelessness, optimism, and suicidal ideation: A study of Ghana and U.S. college samples. *Cross-Cultural Research, 33,* 227–238.

Garmezy, N., Masten, A. S., & Tellegen, A. (1984). The study of stress and competence in children: A building block for developmental psychopathology. *Child Development, 55,* 97–111.

Giaconia, R. M., Reinherz, H. Z., Silverman, A. B., Pakiz, B., Frost, A. K., & Cohen, E. (1995). Traumas and posttraumatic stress disorder in a community population of older adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 34,* 1369–1380.

Green, B. L., Korol, M., Grace, M. C., & Vary, M. G. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 945–951.

Harlacher, T., Okot, F. X., Obonyo, C. A., Balthazard, M., & Atkinson, R. (2006). *Traditional ways of coping in Acholi.* Kampala, Uganda: Caritas.

Herman, J. L. (1992). Complex PTSD: A syndrom in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress, 5,* 377–391.

Hoge, E. A., Austin, E. D., & Pollack, M. H. (2007). Resilience: Research evidence and conceptual considerations for posttraumatic stress disorder. *Depression and Anxiety, 24,* 139–152.

Kanagaratnam, P., Raundalen, M., & Asbjornsen, A. E. (2005). Ideological commitment and posttraumatic stress in former Tamil child soldiers. *Scandinavian Journal of Psychology, 46,* 511–520.

Kazdin, A. E., Kraemer, H. C., Kessler, R. C., Kupfer, D. J., & Offord, D. R. (1997). Contributions of risk-factor research to developmental psychopathology. *Clinical Psychology Review, 17,* 375–406.

King, L. A., King, D. W., Fairbank, J. A., Keane, T. M., & Adams, G. A. (1998). Resilience-recovery factors in posttraumatic stress disorder among female and male Vietnam veterans: Hardiness, postwar social support, and additional stressful life events. *Journal of Personality and Social Psychology, 74,* 420–434.

Kobasa, S. C., Maddi, S. R., & Kahn, S. (1982). Hardiness and health: A prospective study. *Journal of Personality and Social Psychology, 42,* 168–177.

Kohrt, B. A., Jordans, M. J., Tol, W. A., Speckman, R. A., Maharjan, S. M., Worthman, C. M., et al. (2008). Comparison of mental health between former child soldiers and children never conscripted by armed groups in Nepal. *Journal of the American Medical Association, 300,* 691–702.

Kubany, E. S. (1994). A cognitive model of guilt typology in combat-related PTSD. *Journal of Traumatic Stress, 7,* 3–19.

Kubany, E. S., Abueg, F. R., Kilauano, W. L., Manke, F. P., & Kaplan, A. S. (1997). Development and validation of the sources of Trauma-Related Guilt Survey–War-zone version. *Journal of Traumatic Stress, 10,* 235–258.

Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the Trauma-Related Guilt Inventory (TRGI). *Psychological Assessment, 8,* 428–444.

Luthar, S. S. (2006). Resilience in development: A synthesis of research across five decades. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (2nd ed., *Vol. 3,* pp. 739–795). Hoboken, NJ: Wiley.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71,* 543–562.

Macksoud, M. (1992). Assessing war trauma in children: A case study of Lebanese children. *Journal of Refugee Studies, 5,* 1–15.

Marshall, G. N., Orlando, M., Jaycox, L. H., Foy, D. W., & Belzberg, H. (2002). Development and validation of a modified version of the Peritraumatic Dissociative Experiences Questionnaire. *Psychological Assessment, 14,* 123–134.

Masten, A. S. (2001). Ordinary magic. Resilience processes in development. *American Psychologist, 56,* 227–238.

Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19,* 921–930.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments. Lessons from research on successful children. *American Psychologist, 53,* 205–220.

Masten, A. S., Cutuli, J. J., Herbers, J. E., & Reed, M.-G. J. (2009). Resilience in development. In C. R. Snyder & S. J. Lopez (Eds.), *The handbook of positive psychology* (2nd ed., pp. 117–131). New York: Oxford University Press.

McCullough, M. E., Bellah, C. G., Kilpatrick, S. D., & Johnson, J. L. (2001). Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin, 27,* 601–610.

McCullough, M. E., Bono, G., & Root, L. M. (2007). Rumination, emotion, and forgiveness: Three longitudinal studies. *Journal of Personality and Social Psychology, 92,* 490–505.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Jr., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology, 75,* 1586–1603.

McKay, S., Robinson, M., Gonsalves, M., & Worthen, M. (2006). *Girls formerly associated with fighting forces and their children: Returned and neglected.* London: Coalition to Stop the Use of Child Soldiers.

Medeiros, E. (2007). Integrating mental health into post-conflict rehabilitation: The case of Sierra Leonean and Liberian "child soldiers." *Journal of Health Psychology, 12,* 498–504.

Mollica, R. F., McInnes, K., Poole, C., & Tor, S. (1998). Dose-effect relationships of trauma to symptoms of depression and posttraumatic stress disorder among Cambodian survivors of mass violence. *British Journal of Psychiatry, 173,* 482–488.

Mulatu, M. S. (1997). *Stress, coping, and adaptation of Ethiopian adolescents: Testing the applicability of the transactional model of stress and coping.* Unpublished doctoral dissertation, Queen's University, Kingston, ON, Canada.

Naker, D. (2005). *Violence against children: The voices of Ugandan children and adults.* Kampala, Uganda: Save the Children in Uganda.

Office of the Special Representative of the Secretary-General for Children and Armed Conflict. (2006). *Report to the General Assembly. A/61/275.* Retrieved January 15, 2008, from http://www.un.org/children/conflict/english/reports.html

Okello, J., Onen, T. S., & Musisi, S. (2007). Psychiatric disorders among war-abducted and non-abducted adolescents in Gulu district, Uganda: A comparative study. *African Journal of Psychiatry, 10,* 225–231.

Olsson, C. A., Bond, L., Burns, J. M., Vella-Brodrick, D. A., & Sawyer, S. M. (2003). Adolescent resilience: A concept analysis. *Journal of Adolescence, 26,* 1–11.

Onyut, L. P., Neuner, F., Schauer, E., Ertl, V., Odenwald, M., Schauer, M., et al. (2005). Narrative Exposure Therapy as a treatment for child war survivors with posttraumatic stress disorder: Two case reports and a pilot study in an African refugee settlement. *BMC Psychiatry, 5*, 1–9.

Orth, U., Montada, L., & Maercker, A. (2006). Feelings of revenge, retaliation motive, and posttraumatic stress reactions in crime victims. *Journal of Interpersonal Violence, 21*, 229–243.

Oshio, A., Kaneko, H., Nagamine, S., & Nakaya, M. (2003). Construct validity of the Adolescent Resilience Scale. *Psychological Report, 93*, 1217–1222.

Ovuga, E., Boardman, J., & Wasserman, D. (2007). Integrating mental health into primary health care: Local initiatives from Uganda. *World Psychiatry, 6*, 60–61.

Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychological Bulletin, 129*, 52–73.

Pargament, K. I., Smith, B. W., Koenig, H. G., & Perez, L. (1998). Patterns of positive and negative religious coping with major life stressors. *Journal for the Scientific Study of Religion, 37*, 710–724.

Pelcovitz, D., van der Kolk, B., Roth, S., Mandel, F., Kaplan, S., & Resick, P. (1997). Development of a criteria set and a structured interview for disorders of extreme stress (SIDES). *Journal of Traumatic Stress, 10*, 3–16.

Peterson, C. (2000). The future of optimism. *American Psychologist, 55*, 44–55.

Pfefferbaum, B. (1997). Posttraumatic stress disorder in children: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1503–1511.

Pham, P., Vinck, P., & Stover, E. (2007). *Abducted. The Lord's Resistance Army and forced conscription in northern Uganda*. Berkeley, CA: Berkeley-Tulane Initiative on Vulnerable Populations.

Pynoos, R. S., Steinberg, A. M., & Wraith, R. (1995). A developmental model of childhood traumatic stress. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation* (pp. 72–95). Oxford, England: Wiley.

Qouta, S., Punamaki, R. L., Miller, T., & El-Sarraj, E. (2008). Does war beget child aggression? Military violence, gender, age and aggressive behavior in two Palestinian samples. *Aggressive Behavior, 34*, 231–244.

Rescorla, L., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., et al. (2007). Epidemiological comparisons of problems and positive qualities reported by adolescents in 24 countries. *Journal of Consulting and Clinical Psychology, 75*, 351–358.

Rutter, M. (1985). Resilience in the face of adversity. Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry, 147*, 598–611.

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry, 57*, 316–331.

Scheier, M. F., & Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research, 16*, 201–228.

Shaw, J. A. (2003). Children exposed to war/terrorism. *Clinical Child and Family Psychology Review, 6*, 237–246.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl. 20), 22–33.

Taylor, S. E., Kemeny, M. E., Reed, G. M., Bower, J. E., & Gruenewald, T. L. (2000). Psychological resources, positive illusions, and health. *American Psychologist, 55*, 99–109.

Terr, L. C. (1991). Childhood traumas: An outline and overview. *American Journal of Psychiatry, 148*, 10–20.

Thabet, A. A., Abed, Y., & Vostanis, P. (2004). Comorbidity of PTSD and depression among refugee children during war conflict. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45*, 533–542.

Twenge, J. M., & Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort differences on the children's depression inventory: A meta-analysis. *Journal of Abnormal Psychology, 111*, 578–588.

United Nations Children's Fund. (2007). The Paris Principles. Principles and guidelines on children associated with armed forces or armed groups. Retrieved June 10, 2008, from http://www.unhcr.org/cgi-bin/texis/vtx/refworld/rwmain?docid=465198442

van der Kolk, B. A. (2005). Developmental trauma disorder. Towards a rational diagnosis for chronically traumatized children. *Psychiatric Annals, 35*, 401–408.

van der Kolk, B. A. (2007). The developmental impact of childhood trauma. In L. J. Kirmayer, R. Lemelson, & M. Barad (Eds.), *Understanding trauma: Integrating biological, clinical, and cultural perspectives* (pp. 224–241). New York: Cambridge University Press.

Vanderbilt-Adriance, E., & Shaw, D. S. (2008). Conceptualizing and re-evaluating resilience across levels of risk, time, and domains of competence. *Clinical Child and Family Psychology Review, 11*, 30–58.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth*. New York: McGraw-Hill.

Werner, E. E., & Smith, R. S. (2001). *Journeys from childhood to midlife: Risk, resilience and recovery*. Ithaca, NY: Cornell University Press.

Wessells, M. G. (2006). *Child soldiers: From violence to protection*. Cambridge, MA: Harvard University Press.

Yehuda, R., & Flory, J. D. (2007). Differentiating biological correlates of risk, PTSD, and resilience following trauma exposure. *Journal of Traumatic Stress, 20*, 435–447.

Yehuda, R., & LeDoux, J. (2007). Response variation following trauma: A translational neuroscience approach to understanding PTSD. *Neuron, 56*, 19–32.

Yule, W. (2002). Posttraumatic stress disorders. In M. Rutter & E. A. Taylor (Eds.), *Child and adolescent psychiatry* (pp. 520–528). Oxford, England: Blackwell.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1114–1128

The Kenyan Political Conflict and Children's Adjustment

Mumbe Kithakye and
Amanda Sheffield Morris
Oklahoma State University

Andrew M. Terranova
Stephen F. Austin State University

Sonya S. Myers
Vanderbilt University

This study examined pre- and postconflict data from 84 children, ages 3–7 years, living in Kibera, Kenya, during the December 2007 political conflict. Results indicate that children's disaster experiences (home destruction, death of a parent, parent and child harm) are associated with adjustment difficulties and that emotion regulation is an important protective factor postdisaster. Specifically, severity of the disaster experience was associated with increased aggression and decreased prosocial behavior. Emotion regulation was associated with less aggression and more prosocial behavior postconflict. Findings are discussed in the context of a developmental, systems-oriented perspective of the impact of disasters on child adjustment.

While young children do not always present adult-like responses to disaster, exposure to disasters poses serious threats to the psychological and physical health of young children (for a review, see La Greca, Silverman, Vernberg, & Roberts, 2002; La Greca, Silverman, & Wasserstein, 1998; Lazarus & Cohen, 1977; McFarlane, 1987; Onyango, 1998; Ronen, 2002). In Sub-Saharan Africa, children below the age of 15 compose 44% of the population, and millions of these children are exposed to man-made disasters due to civil wars and political conflicts in the region (World Bank, 2007). In the year 2000, for example, 23 civil wars occurred worldwide, 10 of which took place in Sub-Saharan Africa (Bodea & Elbadawi, 2008). Researchers estimate that one in five people are negatively affected by armed conflict in Sub-Saharan Africa (Kimenyi & Ndungu, 2005). Yet, there is little research examining the effects of political conflict on the children's psychosocial adjustment. Drawing from a developmental, systems-oriented perspective (Wapner & Demick, 1998, 2002), the current article contributes to the field of disaster research by examining the impact of the Kenyan political conflict on children's psychosocial functioning, while considering children's preconflict adjustment.

Conflict in Kenya and the Current Study

Prior to 2007, Kenya was considered a nation with a low likelihood of conflict or civil war despite being surrounded by nations long involved in civil war. Sporadic clashes that did occur in Kenya were driven by land disputes, not lootable supplies or political purposes (Collier, Hoeffler, & Sambanis, 2005; Kimenyi & Ndungu, 2005). The political conflict that began shortly after the December 2007 presidential elections brought the nation closer to civil war than any other event since the nation's independence. The two primary candidates for presidency represented the largest tribes in Kenya, the Kikuyus and the Luos. Other major ethnic groups aligned themselves with either group based on historical ties and guidance from political leaders. This polarization led to acts of violence all over the nation, with the worst incidence occurring in the cities of Kisumu and Nairobi, and in Matharae and Kibera (two major slums in Nairobi). The conflict lasted close to 3 months and resulted in an estimated 1,000 deaths and over 100,000 internally

---

Correspondence concerning this article should be addressed to Amanda Sheffield Morris, Department of Human Development and Family Science, Oklahoma State University, Main Hall 1110, Tulsa, OK 74106-0700. Electronic mail may be sent to amanda.morris@okstate.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0007

displaced children. Subsequently, close to 80,000 of these children remained in displacement camps for months after the conflict subsided (Majtenyi, 2008; United Nations Children's Fund [UNICEF], 2008). While most assistance given to families experiencing displacement focused on meeting children's basic needs, little attention was given to children's psychosocial adjustment as a result of these experiences (Kamatsiko, 2006). For this reason, the current study sought to examine changes in children's psychosocial adjustment following political conflict.

Although man-made disasters and natural disasters are generally related to similar psychological difficulties, understudied man-made disasters, such as the political conflict in Kenya, are important to study because political violence has specifically been linked to a variety of negative child outcomes (Sagi-Schwartz, Seginer, & Abdeen, 2008). Man-made and technological disasters have been linked to aspects of children's negative adjustment, such as a weakening sense of control, physical and emotional disturbances, somatic complaints, heightened aggression, fear, withdrawal, and posttraumatic stress symptoms (Davidson, Baum, & Collins, 1982; Gaffney, 2006; Sagi-Schwartz et al., 2008). The Kenyan political conflict occurred with little warning and lasted 3 months without a definitive ending point, characteristic of disasters which have been found to enhance negative consequences already associated with exposure (Baum, Fleming, & Davidson, 1983; Bolin, 1985; Davidson & Baum, 1986; Fritz & Marks, 1954; Lifton & Olsen, 1976).

Developmental, Systems-Oriented Perspective and Resilience

The developmental, systems-oriented perspective offers an ideal venue for examining the effects of disaster-related events, such as the Kenyan political conflict, on child functioning because its approach accounts for dynamic disruptions in person-in-environment systems that affect later development (Wapner & Demick, 1998). Broadly, the developmental, systems-oriented perspective asserts that individuals develop within an environmental context and that both the individual and environment interact at three levels of operation: physical, intrapersonal or interpersonal, and sociocultural levels (Demick & Wapner, 1988; Wapner & Demick, 1998). Whereas the physical level of the individual involves biological functioning and factors such as age, illness, and health, the physical level of the environment involves objects such as home destruction caused by disasters and relocation. The intrapersonal level of the individual encompasses psychological functioning such as attitudes, emotionality, and regulatory abilities, and the analogous level of the environment involves interpersonal factors such as peer and family relations. Sociocultural factors within the individual include factors such as gender, ethnicity, and one's role in society, and the sociocultural factors within the environment involve factors such as school and political transitions (Wapner & Demick, 1998).

Disasters (e.g., floods, hurricanes, terrorism, political conflicts) are thought to disrupt functioning at all levels of the individual and environmental systems (Wapner, 1987; Wapner & Demick, 1998), and these disruptions have been linked to psychosocial adjustment difficulties in children (Jones & Kafetsios, 2005; Moss et al., 2007; Qouta, Punamäki, Miller, & El-Sarraj, 2008; Rousseau, Drapeau, & Platt, 1999). However, not all children are negatively affected by disasters and political conflicts. This ability to positively adapt or mitigate the maladaptive reactions in the face of adversity is often labeled *resilience* (Luthar, Cicchetti, & Becker, 2000; Masten, 1994, 2007; Masten, Best, & Garmezy, 1990). Recent advances in the conceptualization and study of resilience suggest that the process of resilience can operate at and across multiple levels within dynamic systems (Masten, 2007). Thus, just as disasters can negatively impact individuals and their environments at physical, intrapersonal/interpersonal, and social levels, factors associated with these domains can also promote the process of resilience.

Physical Impact

Disasters, such as the Kenyan conflict, likely impact the physical level of the individual by causing physical harm and impact the physical environment by the destruction of homes, buildings, and property. Additionally, factors at the physical level, such as age, have been related to adjustment following disasters. A significant amount of research indicates that there is a dose–response relation between disaster exposure and adjustment (Klingman, 2002; La Greca & Prinstein, 2002; Lengua, Long, Smith, & Meltzoff, 2005; Moore & Moore, 1996; Silverman & La Greca, 2002; Yule & Williams, 1990), with direct and more severe disaster exposure (e.g., physical harm) leading to more adjustment difficulties. A critical review by Pfefferbaum et al. (2003) highlighted the importance of recognizing differences in the severity of exposure to disasters such as September 11, the Oklahoma City bombing, and the 1998 American embassy bombing in Nairobi, Kenya. Children directly exposed to the tragedies evidenced more trauma than those indirectly exposed.

Findings regarding age-related differences in reactions to disaster have been mixed but generally indicate that disasters negatively affect people of all developmental levels (for a review, see Bell, Greene, Fisher, & Baum, 2005; Silverman & La Greca, 2002). Some of the nonsignificant findings regarding developmental differences are likely the result of examining potential developmental differences in a narrow age range of participants, such as only focusing on children in late childhood (Vernberg, La Greca, Silverman, & Prinstein, 1996). One of the studies examining the broadest developmental range suggested the impact of disasters was greatest during middle adulthood (Thompson, Norris, & Hanacek, 1993). When focusing on the periods of childhood and adolescence, a curvilinear, inverted U-shaped association between natural disaster exposure and adjustment difficulties was discovered with adjustment difficulties increasing throughout childhood up until early adolescence and then decreasing into late adolescence (McDermott & Palmer, 2002). Most conflict researchers appear to consider narrow age ranges as well, emphasizing age-relevant impacts of conflict, such as child mortality among children below the age of 5, dropping out of school among school-age children, or sexual exploitation among adolescents (Barbara, 2000; Honwana, 2006).

Another feature of disasters, whether natural or man-made, is destruction of the home. Researchers are increasingly acknowledging the importance of home destruction to disaster victims. Findings indicate that home destruction negatively impacts individuals exposed to disaster, with higher levels of posttraumatic stress disorder (PTSD) symptoms and depression evidenced among those who experience severe or total home destruction (Gerrity & Steinglass, 2003; Goto, Wilson, Kahana, & Slane, 2006). Few studies have examined the effects of home destruction on children. An exception is Houlihan, Ries, Polusny, and Hanson (2008) who examined the impact of home relocation following a major tornado on children in Grades 3–12. Children in Grades 3–6 who relocated, due to severe home destruction, presented higher levels of internalizing behavior problems. Children in Grades 7–12 who relocated presented more externalizing behavior problems (Houlihan et al., 2008). Another study found that children whose homes were destroyed by a wildfire were at greater risk of posttraumatic stress symptoms (Jones, Ribbe, & Cunningham, 1994). More research is needed on child functioning following home destruction, particularly during political conflicts and war.

Intrapersonal and Interpersonal Factors

Disasters can also negatively impact children's socioemotional functioning by disrupting children's interpersonal interactions with their families because of parental harm, parental death, parental separation, or parental psychopathology. Additionally, intrapersonal factors such as children's abilities to manage their negative reactions likely mitigate some of the effects of disasters and promote resilience. Previous disaster research suggests that there are key adaptive systems that can affect outcomes of children facing threats, including attachment, social support within families and communities, and self-regulatory abilities. Self-regulation is particularly important to examine in relation to disaster events because being able to adapt and function in any situation requires aspects of regulatory control, such as determining/inhibiting actions, controlling emotions, and being able to focus attention in highly arousing situations (Masten, 2004). Indeed, self-regulatory abilities have been one of the more prominent factors investigated in the study of resilience and have been associated with more positive adaptation among children both during and following traumatic events (Eisenberg, Smith, Sadovsky, & Spinrad, 2004; Masten, 2007). Research suggests that children who cope poorly with disaster exposure and the resulting emotional distress suffer more severe stress reactions. Alternatively, children who effectively regulate their emotions are more resilient (Almqvist & Broberg, 1999; Kocijan-Hercigonja, Rijavec, Marusic, & Hercigonja, 1998; La Greca, Silverman, Vernberg, & Prinstein, 1996; La Greca et al., 1998; Shannon, Lonigan, Finch, & Taylor, 1994; Vernberg et al., 1996; Weems et al., 2007). It is likely that aspects of voluntary (i.e., effortful) self-control serve as protective factors in response to disaster exposure (Masten & Obradović, 2008). Despite the interest in the relation between self-regulation and resilience, there is little research on emotion regulation and adjustment following a disaster.

Emotion regulation involves components of temperament (e.g., affect, activity, and attention) typically defined in terms of the effortful processes that afford young children the ability to inhibit dominant responses and activate subdominant responses in times of stress (Eisenberg, Hofer, & Vaughan, 2007; Rothbart & Bates, 2006). One example includes inhibiting flight reactions and taking time to consider other responses, particularly in fear-arousing situations. In response to a traumatic event, emotion regulatory abilities allow children to shift their attention from distressing aspects of the event and inhibit maladaptive responses while still being able to process information related to the event (Lengua et al., 2005). In adults, improvements in regulatory abilities are related to the reduction in emotional and behavioral difficulties following exposure to war (Cloitre, Karestan, Cohen, & Han, 2002; Cloitre, Stovall-McClough, Miranda, & Chemtob, 2004), and emotion regulatory abilities protect children from long-term PTSD symptoms following disasters (Terranova, Boxer, & Morris, 2009).

Additionally, emotion regulation generally is associated with positive adjustment in young children because it enables them to react to specific environmental demands, or in the current context disaster experiences, with more positive adaptive responses (Eisenberg & Morris, 2002; Eisenberg et al., 2004; Eisenberg et al., 2007). Thus, more developed emotion regulatory abilities likely serve as a protective factor, making children more resilient to the adjustment difficulties commonly following exposure to disasters and political violence. It is important to note, however, that these constructs have been primarily studied in Western societies. Nonetheless, several conflict researchers have noted the possible role of children's regulation in coping, resilience, and functioning in times of war (Kostelny, 2006; Machel, 1996).

Intrapersonal characteristics contributing to positive adjustment following a disaster are often present prior to the disaster. However, due to the sudden and unexpected nature of most disasters, researchers rarely have the opportunity to examine the influence of predisaster intrapersonal characteristics on functioning. La Greca et al. (1998) reported high attention problems and low academic achievement prior to a disaster as risk factors for postdisaster stress symptoms. Other studies have found that children with more psychological difficulties prior to the disaster presented more symptoms of trauma, less coping skills and more behavioral problems following the disaster (McFarlane, 1987; Silverman & La Greca, 2002). Examining the relation between pre- and postdisaster adjustment may assist in identifying groups of children who are more likely to present poor adjustment following a conflict. Furthermore, the inclusion of predisaster adjustment allows for increased understanding of interpersonal factors associated with positive adjustment or the process of resilience postconflict.

Social support is another important interpersonal mechanism through which young children are impacted by trauma, with greater social support minimizing the negative effects of disaster exposure (e.g., Fleming, Baum, Gisriel, & Gatchel, 1987; Wapner & Demick, 2002). When parental health and functioning are negatively affected by a traumatic event, the support that parents are able to provide is limited. Child functioning is also likely affected as young children look to their parents to learn how to react to difficult or traumatic circumstances (A. S. Morris, Silk, Steinberg, Myers, & Robinson, 2007). Separation from a parent due to a disaster, parental harm, or even parental death are especially detrimental to children. Children who are separated from their parents experience more emotional distress during a natural disaster than those who are not separated (McFarlane, 1987; Silverman & La Greca, 2002). Similar findings are reported among children affected by war. Increases in aggression were reported among Israeli boys who experienced the death of their father due to war (Levy-Shiff, 1982). In South Africa, preschoolers displayed a variety of responses following detention of a parent for political reasons. Some children appeared more irritable, moody, and clingy during the detention, while others presented stomach aches and nightmares (Barbara, 2000). Particularly among young children, it is important to consider the ways in which family members, specifically parents, are impacted by natural and man-made disasters.

Sociocultural Impacts

A growing body of research also indicates that variations in postdisaster functioning are partially attributable to sociocultural factors, which include gender, socioeconomic status (SES), and ethnicity or culture (for a review, see Peek & Mileti, 2002; Silverman & La Greca, 2002). While the majority of the research supporting this statement is taken from natural disaster research (e.g., research about hurricanes, fires, or earthquakes), similar patterns are expected following political conflicts and other man-made disasters. Regarding the demographic characteristics of children, natural disaster research is sometimes mixed; however, some conclusions can be drawn. Generally, being female (e.g., Smith, Perrin, Yule, Hacam, & Stuvland, 2002; Vernberg et al., 1996; Weems et al., 2007) and of minority status (Brockerhoff & Hewitt, 2000; La Greca et al., 1996; Rabalais, Ruggiero, & Scotti, 2002; Shannon et al., 1994) are associated with more adjustment difficulties. Gender differences may be related to different propensities for psychopathology among males and females, but more research is needed to fully understand these differences, particularly among young children. Because males and females often differentially experience atrocities due to gender-based violence and sexual exploitation, gender is likely to be associated with differences in the experience of conflict. In the current study, child sex is examined as an indicator of child gender.

Disaster research also suggests that lower SES and related factors place minorities at increased risk of psychosocial difficulties (Bonanno, Galea, & Bucciarelli, 2007; Miller & Heldring, 2004). In the developing countries of Sub-Saharan Africa, the term *minority* is rarely used; nonetheless, the notion that groups of persons who represent a smaller proportion of the national population is relevant. Each of the 47 nations in Sub-Saharan Africa is home to many (∼ 30–500) different people groups or tribes. Due to ethnic diversity, nations are made of some groups that can be identified as majority groups and others as minority groups, with reference to numbers of people identifying with a specific ethnic identity. Majority and minority status, however, does not always reflect differences in SES or access to resources (Collier et al., 2005). In Kenya, ethnic minority status is often defined as either an "ethnic, linguistic, or political group" who lacks a dominant position, a group determined to preserve its ethnic identity in response to the larger society, or a group likely to be discriminated against due to their language, ethnicity, or religion (Makoloo, 2005). For example, groups may consider themselves in the minority due to the political dominance of one group at the expense of another (e.g., control of state, disbursement of favors to specific communities), religion-based persecution (e.g., discrimination and denial of rights), or on the basis of their way of life (e.g., pastoral vs. hunter-gatherer communities). Therefore, conflict difficulties may have less to do with ethnic group, or cultural identity and more to do with access to resources and location (Liddell, 1994). In the current study, we were not able to formally examine effects by tribal group as individual child group membership information was not collected. However, open-ended teacher responses on the impact of ethnicity or tribal status allowed some insight into the role of tribal group membership.

Current Study

Four months prior to the 2007 political conflict in Kibera, Kenya, data were gathered on children's socioemotional adjustment. Shortly after the end of the conflict, 3 months after the December 2007 elections, data were collected on the same children's postconflict allowing for a rare examination of the impact of politically related conflict on child functioning in Sub-Saharan Africa. Within this context, the current investigation utilized a longitudinal design to examine the impact of child and parental disaster exposure on children's postdisaster functioning (emotional symptoms, aggression, and prosocial behavior), while controlling for children's predisaster functioning. There is minimal research on children's psychosocial functioning following political conflicts in Sub-Saharan Africa, and most research has focused on physical health and mortality. The current study extends extant findings by examining changes in children's socioemotional adjustment (e.g., prosocial behaviors and internalizing and externalizing problems) following the Kenyan conflict. As has been reported in previous studies, a dose–response relation between disaster exposure and adjustment difficulties was expected. Specifically, we expected increases in emotional symptoms and aggression, and decreases in prosocial behavior, particularly among children who were more negatively impacted by the political conflict. Given the protective role of emotion regulation during adversity, it was expected that emotion regulation would act as a moderator between disaster experience and adjustment, such that children with better regulatory abilities would display fewer adjustment problems when exposed to disaster. Patterns of findings were also examined for sex and age differences.

Method

Participants

Pre- and postconflict data were collected on 84 children (46 female and 38 male) from their teachers ($n = 2$). Similar to the general ethnic breakdown in Kibera, participating children represented a broad range of ethnic backgrounds, including Kikuyus, Kambas, Luos, and Luhyas. At the start of the study, the mean age was 4.9 years ($SD = 1.23$) with 11% of the children identified as 3-year-olds, 27% as 4-year-olds, 21% as 5-year-olds, 20% as 6-year-olds, and 10% as 7-year-olds. All of the children resided in Kibera, one of the largest slums in Africa. Kibera is a densely populated area (estimated 800,000 people per 2 km^2) located in the capital city, Nairobi, of Kenya. Kibera is an informal urban settlement, characterized by high levels of poverty, inadequate access to safe water and sanitation, poor quality housing, and high levels of HIV/AIDS (United Nations Human Settlements Program, 2003). Residents of Kibera experienced especially high levels of postelection conflict in January and February 2008.

Measures

Severity of conflict experience. Due to the physical and logistical challenges of gathering data in Kibera, disaster severity data were collected only from teachers. As the teachers resided in the densely populated, close-knit community from which study participants resided, the teachers were well informed about children's conflict exposure and spoke with the children about their experiences. Teachers were asked to respond to several open-ended questions regarding the overall effects of the conflict on children and also provided responses to specific conflict exposure items for each child in order to asses children's individual level of exposure. The relevant items of conflict exposure included whether the child ''was separated from parent,'' ''was physically harmed,'' ''had a home destroyed,'' ''had a parent die,'' or ''had a parent physically harmed.'' Responses were coded a 1 if the child experienced the event and a 0 if the child did not experience the event. A disaster severity variable was created to reflect overall conflict experience by summing the values of the responses on the five disaster items presented above.

Child adjustment pre- and postconflict. Teachers reported on children's pre- and postconflict adjustment by completing the aggressive behavior, prosocial behavior, and anxious-fearful (emotional symptoms) scales of the Child Behavior Scale (CBS; Ladd & Profilet, 1996). Teachers responded to each item on a 3-point Likert scale ranging from *doesn't apply* to *certainly applies*. The scales have demonstrated good internal consistency (αs = .77–.96) and construct validity, particularly among children of different ethnicities and SES groups (Ladd & Profilet, 1996). In the current study, *aggression* was assessed via the aggressive behavior scale (seven items, $\alpha = .80$) using items such as ''kicks, bites, hits'' and ''taunts, teases.'' *Emotional symptoms* were assessed via the anxious-fearful scale (four items, $\alpha = .45$) with sample items such as ''is worried'' and ''cries easily.'' *Prosocial behavior* was assessed via the prosocial scale (six items, $\alpha = .81$) using items such as ''helps others'' and ''cooperative with peers.''

Child characteristics. Child age and sex were reported by parents and teachers prior to the conflict. Reports on children's *emotion regulation* were gathered postconflict via the effortful control scale of the Child Behavior Questionnaire–Shortened Scales (CBQ; Fabes, 1994; Goldsmith & Rothbart, 1991), which assesses a temperamental component of children's emotion regulation (Eisenberg & Morris, 2002). Teachers were asked to rate how true an item was for the child on a 7-point Likert scale (1 = *extremely untrue* to 7 = *extremely true*). Internal consistency estimates of the CBQ range from .67 to .94 (Goldsmith & Rothbart, 1991; see Fabes, 1994, for the reliability of the shortened scales), and scales from the CBQ have been correlated with similar observed constructs and child adjustment (Eisenberg et al., 1997). The effortful control scale is created by combining the attention focusing, attention shifting, and inhibitory control scales in an approach that has been used in prior research (e.g., Morris et al., 2002). Mean scores were created for each scale where higher scores indicate greater emotion regulation. The scores were then standardized and all three scales were summed creating the *emotion regulation* scale. The current study presented adequate reliability estimates for the combined scale ($\alpha = .66$ with 18 items). Sample items on this scale include ''is able to resist laughing or smiling when it isn't appropriate,'' can easily quit working on a project if asked,'' and ''is easily distracted when listening to a story'' (reversed).

Procedure

Children were recruited from a nursery school/day care center developed to serve the needs of the most vulnerable children either due to parental HIV/AIDS infection or death or due to extended periods of unemployment. Prior to the conflict, approximately 100 children attended the nursery school, 92 participated in Time 1 data collection, and 84 children participated in Time 2 data collection following the conflict. Data were collected through questionnaires completed by teachers at both Time 1 and Time 2. Teacher questionnaires were in English as the teachers are fluent English speakers. Preconflict data were collected by research assistants in Kenya, and postconflict data via mail, with directions and consent procedures also completed over the phone. For both pre- and postconflict data collection, Institutional Review Board approval was granted prior to data collection.

Results

Correlations and regression analyses were utilized to examine specific hypotheses, as outlined next. In addition to empirically testing hypothesized relations, teachers' responses to open-ended questions regarding the political conflict and the effect of the conflict on children's adjustment illustrate the role of ethnicity and describe the extent to which children were affected.

Descriptive Findings

Responses to specific exposure questions indicated variability in the disaster experience across the sample. Calculated percentages for the disaster variables indicate that: 43% had their home destroyed, 20% were physically harmed, 11% were separated from their parents, 13% had a parent die, and 26% had a parent physically harmed. Parent death was not included in parent harm or parent separation codes to ensure more independence among these variables. Thus, the percentages presented above reflect the percentage of individual children who experienced parental death, parental physical harm without counting harm leading to death, and parental separation not occurring as a result of parental death. Disaster severity was calculated a composite of the five disaster variables (coded 0 or 1), with a mean severity exposure score of 1.2 (see Table 1) in our sample and scores ranging from 0 to 4.

Teachers' responses to open-ended questions identified additional negative impacts of the conflict on children beyond those we had identified. Teacher 1 noted, "Several houses were torched and so children lack basic needs such as food, shelter, and clothing." Teacher 2 reported, "Some children were sexually harassed," and "Some children lost their close relatives." Teachers also observed increases in children's fear and negativity toward children in other tribal groups stating, "Most children have developed negative attitudes towards others and hence group themselves in tribal lines." No quantitative statistical analyses were performed using these data, as the open-ended questions assessed aggregate child functioning and experiences, intended solely for descriptive purposes.

Impact of Disaster Experience on Adjustment

We expected children's negative experiences during the conflict to be significantly associated with children's adjustment difficulties. In order to test this hypothesis, we examined correlations among disaster variables and adjustment postdisaster (emotional symptoms, aggression, and prosocial behavior). Child predisaster and postdisaster adjustment were not correlated, and predisaster adjustment was not associated with disaster experience variables; as such, the predisaster adjustment variables (emotional symptoms, aggression, and prosocial behavior) were not included in the correlation table (see Table 1). Regression analyses were conducted to examine the impact of disaster variables on postdisaster adjustment, controlling for predisaster adjustment, sex, and age of the child in each model. Although predisaster adjustment, sex,

Table 1
Correlations and Descriptive Statistics for Selected Study Variables

Variables	1	2	3	4	5	6	7	8	9	10
1. Disaster severity	–									
2. Home destroyed	.64***	–								
3. Parent separation	.50***	.04	–							
4. Parent death	.33**	.33**	−.05	–						
5. Parent harmed	.81***	.30*	.20	−.14	–					
6. Child harmed	.69***	.06	.27*	.06	.63***	–				
7. Emot. regulation^a	−.49***	−.32**	−.24*	−.16	−.49***	−.24*	–			
8. Emot. symptoms^a	.05	.05	.12	.28**	−.18	.05	−.12	–		
9. Aggression^a	.25*	.27*	.07	.13	.14	.09	−.62***	.05	–	
10. Prosocial behav.^{a}	−.44***	−.31***	−.13	−.21*	−.35***	−.28*	.60***	−.10	−.54***	–
M	1.16	.46	.13	.13	.30	.21	.00	2.05	1.70	2.35
SD	1.11	.50	.33	.34	.46	.41	2.33	.39	.42	.46

Note. Emot. = emotion; Behav. = behavior; Emot. Regulation = sum of z-scores.
^{a}Postdisaster.
*$p < .05$. **$p < .01$. ***$p < .001$.

and age were not significantly correlated with postdisaster adjustment or disaster variables, they were included in the regression analyses in order to control for predisaster adjustment. Separate models were computed for each adjustment and disaster variable, and for the composite disaster variable.

As expected, disaster variables were correlated with adjustment in expected directions, particularly prosocial behavior (see Table 1), with the exception of parental separation. Parental death was associated with greater emotional symptoms. Destruction of the home and overall disaster severity were associated with more aggression. Regression analyses examined changes in adjustment following the disaster experience. Regression models examined the contribution of each disaster variable (home destruction, parental death, parental separation, parent harm, child harm, and overall disaster severity) to child adjustment (emotional symptoms, aggression, and prosocial behavior) while controlling for child age, sex and the relevant predisaster adjustment variable (Time 1 emotional symptoms, aggression and prosocial behavior). All significant results reported in the current article represent an alpha level of .05 or below.

Unexpectedly, none of the disaster variables were significantly associated with emotional symptoms, controlling for predisaster emotional symptoms, age, and sex. For aggression, disaster severity, and home destruction, each predicted a significant proportion of variance in their respective postconflict aggression models. Home destruction, child sex, age, and preconflict aggression accounted for 17% of the variance in a significant model of postconflict aggression, $R^2 = .17$, $F(4, 73) = 3.78$. However, home destruction was the only significant predictor in the model, $\beta = .29$. Disaster severity, child sex, age, and preconflict aggression accounted for 17% of variance in a significant model of postconflict aggression, $R^2 = .17$, $F(4, 73) = 3.67$. Disaster severity was the only significant predictor in the model, $\beta = .28$.

For prosocial behavior, each disaster variable significantly predicted a proportion of variance in their respective postconflict prosocial behavior models. Home destruction, child sex, age and preconflict prosocial behavior accounted for 20% of the variance in a significant model of postconflict prosocial behavior, $R^2 = .20$, $F(4, 73) = 4.47$, with sex (0 = *male*, 1 = *female*), $\beta = .25$, and home destruction, $\beta = -.31$, as the only significant predictors in this model. Parent harm, child sex, age and preconflict prosocial behavior accounted for 24% of the variance in a significant postconflict prosocial behavior model, $R^2 = .24$, $F(4, 63) = 5.02$. Sex, $\beta = .33$, and parent harm, $\beta = -.38$, were both significant predictors in the model. Child harm, child sex, age, and preconflict prosocial behavior accounted for 19% of the variance in a significant postconflict prosocial behavior model, $R^2 = .19$, $F(4, 69) = 3.96$, with sex, $\beta = .27$, and child harm, $\beta = -.30$, as the only significant predictors. Parent death, child sex, age, and preconflict prosocial behavior accounted for 12% of the variance in a significant model of postconflict prosocial behavior, $R^2 = .12$, $F(4, 73) = 2.88$. Sex of the child was a significant predictor, $\beta = .24$, and parent death was a marginally significant predictor $\beta = -.20$. Disaster severity, child sex, age and preconflict prosocial behavior accounted for 30% of the variance in a model of postconflict prosocial behavior, $R^2 = .30$, $F(4, 73) = 7.90$. Sex of the child, $\beta = .25$, and disaster severity, $\beta = -.45$, were the only significant predictors in the model.

Impact of Child Characteristics on Adjustment

We hypothesized that children's emotion regulation would be associated with better adjustment postconflict and that emotion regulation would act as a moderator between disaster experience and adjustment, with poor emotion regulation exacerbating the effects of disaster severity. Potential moderating and direct effects of children's age and sex were explored as well. Moderation was tested according to Aiken and West's (1991) guidelines, and Holmbeck's (2002) procedures were used for probing significant interactions.

In both correlation and regression analyses, we found that emotion regulation was associated with postdisaster functioning. Specifically, emotion regulation was associated with less aggression and more prosocial behavior (see Table 1) even when controlling for participant sex, age, prosocial behavior preconflict, and disaster severity (see Table 2). Emotion regulation was not associated with emotional symptoms in correlation or regression analyses. Associations between disaster experiences and emotion regulation are also noted in Table 1, indicating poorer emotion regulation among children with more severe disaster experiences. As expected, emotion regulation moderated the link between disaster severity and prosocial behavior (see Table 3), but moderation was not found for aggression or emotional symptoms. Despite a marginal interaction, significant differences were found among children high and low in emotion regulation. Prosocial behavior was similar among children

Table 2

Regression Analyses Predicting Adjustment From Disaster Experience and Emotion Regulation, Controlling for Predisaster Adjustment, Sex, and Age

	Aggression		Prosocial behavior	
	B	β	B	β
Sex of child	−.12	−.13	.20*	.21*
Age	.05	.14	−.03	−.07
T1 adjustment[a]	.11	.10	.21	.18
Disaster severity	−.02	−.04	−.09*	−.21*
Emotion regulation	−.11***	−.62***	.09***	.47***
R^2	.44		.46	
F	$F(5, 72) = 11.16$***		$F(5, 72) = 12.33$***	

Note. Time 1 (T1) aggression was used as a control when predicting Time 2 (T2) aggression. T1 prosocial behavior was used as a control when predicting T2 prosocial behavior. B is the unstandardized coefficient and β is the standardized coefficient. *Males* = 0, *females* = 1.

[a] Aggression or prosocial behavior.

*$p < .05$. ***$p < .001$.

Table 3

Regression Analyses Predicting Prosocial Behavior From Disaster Experience, Emotion Regulation, and Their Interaction

	Prosocial behavior	
	B	β
Disaster severity	−.06	−.16
Emotion regulation	.06*	.33*
Disaster Severity × Emotion Regulation	.03†	.25†
R^2	.41	
F	$F(3, 79) = 2.37$***	
High emotion regulation	.02	.04
Low emotion regulation	−.14**	−.35**

Note. B is the unstandardized coefficient and β is the standardized coefficient.

$^{\dagger}p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

high in emotion regulation regardless of disaster severity. Among children low in emotion regulation, disaster severity was associated with less prosocial behavior (see Table 3).

In terms of age and sex differences, age was not correlated with any adjustment variables. Sex of the child was correlated with postconflict prosocial behavior and aggression. Independent samples t tests indicated a significant difference in prosocial behavior, $t(82) = 2.60$, and aggression, $t(82) = 2.14$. Teachers reported more prosocial behavior and less aggression in girls compared with boys. Interaction terms were computed separately for sex and age and disaster severity (Sex × Disaster Severity; Age × Disaster Severity) in predicting adjustment. Neither interaction term predicted postdisaster adjustment, suggesting that differences in prosocial behavior and aggression were not influenced by an interaction between the sex or age of the child and overall disaster experiences. Nonetheless, child sex was a significant contributor to variance in child prosocial behavior, although the impact of disaster variables is similar among girls and boys.

Discussion

The paucity of research on children's postconflict adjustment in Sub-Saharan Africa is notable as the numbers of conflicts in the area continue to increase (Collier, 2006). As aid organizations search for ways to best assist children affected by conflict, the current investigation identifies person–environment variables that place children at increased vulnerability for poor functioning, indicating possible points for intervention. We found that parental death was associated with more emotional symptoms. Home destruction, disaster severity, and poor emotion regulation were associated with more aggression, even when controlling for predisaster aggression. We found strong links between disaster experiences and decreases in prosocial behavior as well. Home destruction, parent and child harm, parental death, poor emotion regulation, and overall disaster severity were associated with decreased prosocial behavior. Thus, our results support the premise that environmental level factors (e.g., home destruction) affect the child's ability to function following a disaster, and that the effect of the disaster experience on child functioning can be influenced by person level factors (e.g., emotion regulation).

Teachers' responses to open-ended items indicated changes in children's overall behavior (e.g., loss of appetite, fear) and their social behaviors in direct response to the conflict as well. For example, as reported above, one teacher described children as classifying and relating to one another based on tribe. This behavior is similar to observed adult behaviors during the conflict. Media reports of the conflict detailed attacks based on tribal lines (Majtenyi, 2008), and children were likely exposed to similar statements from family members. Children's aversion to sitting next to classmates from different tribes likely reflected the fear and prejudice adults were expressing. Thus, these findings suggest that increased awareness of tribal groups was a relevant aspect of the postconflict environment, although specific tribal membership was not examined.

Findings in the current study indicate that children who had more severe disaster experiences evidenced increases in aggression and decreases in prosocial behavior. These results are in agreement with other disaster studies that report higher levels of emotional problems and aggression among children exposed to a disaster (McFarlane, 1987; Silverman & La Greca, 2002; Wasserstein & La Greca, 1998; Weems et al., 2007; Yorbik, Akbiyik, Kirmizigul, & Sohmen, 2004). Findings also indicate that emotion regulation is associated with more prosocial behavior and less aggression postconflict. Despite little research into emotion regulation postdisaster among children, these findings concur with expectations of other researchers. For example, Lengua et al. (2005) suggested that children who had lower emotion regulation would show more adjustment difficulties following a disaster. Our findings indicate that emotion regulation abilities are associated with increased prosocial behavior and decreased aggression postconflict, even when controlling for preadjustment levels.

This study examined both child level and environmental factors that have been identified by researchers as potentially relevant to children's postdisaster functioning. The developmental, systems-oriented perspective asserts that aspects within the individual and aspects of the environment interact at multiple levels, and disasters can impact functioning at and across these multiple levels of functioning (Demick & Wapner, 1988; Wapner & Demick, 1998). Additionally, when studying disasters more specifically, Silverman and La Greca (2002) identified individual and environmental factors that play important roles regarding children's adjustment following disasters. Consistent with these perspectives, our findings suggest that child level factors interact with the context of conflict, in some cases exacerbating the impact of the disaster. We found that among children with poor emotion regulation abilities, disaster severity was associated with less prosocial behavior, whereas among children high in emotion regulation, disaster severity was not associated with prosocial behavior. These findings are consistent with recent research regarding the protective role of emotion regulation following natural disasters (Terranova et al., 2009). Recent conceptualizations of resilience include regulatory abilities as well as other protective factors interacting within a dynamic systems framework (Masten, 2007). Thus, subsequent research into adjustment following disasters should also examine potential protective factors from other domains, such as biological, family, community, and cultural factors.

J. Morris, van Ommeren, and colleagues recently acknowledged the need for empirical support of the child-relevant standards on mental and social health interventions in the most widely used humanitarian aid document, the Humanitarian Charter and Minimum Standards in Disaster Response (J. Morris, van Ommeren, Belfer, Saxena, and Saraceno, 2007; Sphere Project, 2004). They noted that minimal effort had been made to base child-related recommendations on empirical or theoretical findings. J. Morris and colleagues identified an immense need for further research examining the relevance of the recommendations for children, suggesting that empirically based research would contribute significantly to the ability of Sphere standards to present efficacious and beneficial intervention recommendations to children encountering disasters worldwide (J. Morris et al., 2007). As such, our examination of factors associated with children's adjustment postconflict may serve as a necessary step in the process of identifying opportunities for interventions that can be child focused, empirically supported, and universally applicable. For example, our findings support programs targeted at improving children's emotion regulation and increasing prosocial behavior postconflict.

There are several limitations in the current study that should be noted. A significant challenge involved choosing measures to assess children's socioemotional adjustment and conflict exposure. Although the CBS has been used in research in many different cultural and ethnic settings (Ladd & Profilet, 1996), research is lacking with children in particularly low-resource settings such as the informal human settlement of Kibera. However, high levels of associations between the scales in expected directions provide some validity for the use of the measure. The only variable that did not align with expected associations was emotional symptoms. As other studies have found significant associations between disaster exposure and preschool children's emotional symptoms (La Greca et al., 2002; McFarlane, 1987; Moss et al., 2007), the lack of findings for emotional symptoms in the current study is likely related to the low reliability of the emotional symptoms scale in the sample ($\alpha = .45$). Low reliabilities may be due to cultural differences in how behavior is described and what types of behaviors are socially acceptable. Generally however, associations between disaster experiences, child characteristics, and child outcomes all presented meaningful patterns, indicating construct validity despite the lack of prior support for use of measures. The items used to measure conflict were chosen to reflect disaster experiences researchers have identified as relevant to postdisaster functioning. By only allowing teachers to indicate whether the children experienced the different aspects of the conflict, we limited the expression of variability in children's disaster experiences. For example, the type of harm a parent experienced was not defined in the item asking whether the child's parent had experienced harm. The term *harm* could be interpreted in many different ways. We also were not able to examine whether the length of disaster experience or separation from a parent influenced child adjustment, as information on the length of separation was not collected. The combination of the exposure items to create a composite disaster severity variable allowed for some variability in disaster to be assessed. Study findings indicate that different aspects of the disaster may affect children in different ways. As such, further research could benefit from more specific descriptions of the disaster experiences.

Another limitation of this study is that teachers served as the only source of data in a restricted age group. With teachers as the only source of data postdisaster, detailed family-level information post-conflict was not available. This limited the ability to examine possible interactions between the child and family level factors. In addition, the only post-conflict family variables assessed were parental separation, death, and harm. Contrary to other disaster studies, parental separation was not associated with child functioning postconflict. A lack of findings may be due to the small numbers of children ($n = 9$) who experienced parental separation or the fact that a distinction was made between parental separation, death, and harm. Despite the absence of findings with parental separation, parental harm and death were associated with child post-disaster functioning, illustrating the importance of parent experiences and responses to postdisaster child adjustment. It also should be noted that the disaster variables are not independent. Many of the children experienced multiple disaster experiences (e.g., home destruction *and* death of a parent). Thus, findings for specific disaster experiences should be examined in light of this limitation. We chose to examine specific disaster experiences given little research on political conflict in Africa; however, the composite disaster findings are more robust to this limitation.

Results associated with age should be interpreted with caution in this study as well. While the sample of preschool and early school-age children provides a wider age range than most disaster studies, the low levels of literacy in the area leave some uncertainty in actual dates of birth for the children in the study, which were reported by parents predisaster. Sex differences in overall patterns of results also were not found. However, we did find differences in the levels of prosocial behavior and aggression. Further research should continue to examine the role of sex postdisaster as other studies have found more deleterious effects among females following disaster (Weems et al., 2007). It is possible that our lack of findings is due to a young sample that probably had similar conflict experience regardless of sex.

One important finding that emerged in this study was the decrease in prosocial behavior linked to disaster experiences and disaster severity. This is not surprising given rising social tensions among adults as a result of political and ethnic polarization; however, it is important to note that this tension was found among children. This is one of the first studies to our knowledge to examine prosocial behavior among children experiencing political conflict, as most disaster research has focused on negative functioning (e.g., aggression, PTSD, or internalizing problems). Our findings indicate that prosocial behavior is an important aspect of child functioning postconflict and that interventions should aim to increase prosocial behavior as well as decrease adjustment difficulties. This is in accordance with Machel's (1996) recommendation that supporting well-being may be more culturally appropriate in areas experiencing armed conflict, noting that the traditional therapy approach of having children talk about their experiences may not provide a cross-culturally appropriate means of coping with trauma (Machel, 1996).

This study provides some empirical support for specific intervention strategies following a political conflict. In line with other researchers and current disaster practices, all efforts should be made to protect children from physical harm during conflicts. In addition to physical difficulties, children who experience harm are more likely to present psychological difficulties following a conflict. The recent move to create ''safe spaces'' during a conflict that serve as an isolated, secure area where weapons are prohibited and children can play, may influence not only physical health but also psychological well-being following conflict (J. Morris et al., 2007; Sphere Project, 2004; UNICEF, 2002). Although we did not find a significant relation between parental separation and child adjustment, effects were found for parental death and harm. It is likely that intervening to strengthen parent functioning postdisaster and the protection of parents during conflict will influence child adjustment postdisaster. Finally, this study supports child-level interventions that support children's abilities to interact in appropriate ways and manage their emotional responses. Children who can control their emotional responses appear to adjust better following a disaster.

In this study, the focus is on children's functioning following political conflict. However, changes beyond the child can significantly influence children's functioning and resilience. Ngige (2008) identified five different levels of potential resilience stemming from her experiences working with war-affected children and families in Sudan: individual, extended family, community, national, and international. As the current study illustrates, resilience at the individual level (e.g., emotion regulation, prosocial behavior) is associated with children's functioning following political conflict. It is possible, however, to move beyond this level and inform our understanding of children's resilience more broadly by examining factors from other systems that impact children (Masten, 2007). A growing interest in how resilience is influenced by factors across systems provides future researchers with the ideal climate to examine factors at the family, community, national, and international levels (Masten & Obradović, 2008). Such findings have the potential to impact policy as well as intervention efforts. As one teacher illustrated with a comment written at the end of the survey, "As in the law of Rwanda, Kenya's constitution should force Kenyans not to classify others in terms of tribes, it should be a Kenyan and not Kikuyus, Kambas, or even Luos." Whether such behaviors can be enforced is beyond the scope of this study. Nevertheless, we can argue that in the same ways that adults' actions can harm children, children also can benefit from positive actions and policies undertaken on their behalf.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Almqvist, K., & Broberg, A. G. (1999). Mental health and social adjustment in young refugee children 3½ years after their arrival in Sweden. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 723–730.

Barbara, J. S. (2000). The psychological effects of war on children. In B. S. Levy & V. W. Sidel (Eds.), *War and public health* (pp. 168–185). Washington, DC: American Public Health Association.

Baum, A., Fleming, R., & Davidson, L. M. (1983). Natural disasters and technological catastrophe. *Environment and Behavior, 15*, 333–354.

Bell, P. A., Greene, T. C., Fisher, J. D., & Baum, A. (2005). *Environmental psychology* (5th ed.). New York: Harcourt.

Bodea, C., & Elbadawi, I. A. (2008). Political violence and underdevelopment. *Journal of African Economics, 17*, ii50–ii96.

Bolin, R. (1985). Disaster characteristics and psychosocial impacts. In B. J. Sowder (Ed.), *Disasters and mental health: Selected contemporary perspectives* (pp. 3–28). Rockville, MD: United States Department of Health and Human Services.

Bonanno, G. A., Galea, S., & Bucciarelli, A. (2007). What predicts psychological resilience after disaster? The role of demographics, resources and life stress. *Journal of Consulting and Clinical Psychology, 75*, 671–682.

Brockerhoff, M., & Hewitt, P. (2000). Inequality of child mortality among ethnic groups in Sub-Saharan Africa. *Bulletin of the World Health Organization, 78*, 30–41.

Cloitre, M., Karestan, K. C., Cohen, L. R., & Han, H. (2002). Skills training in affective and interpersonal regulation followed by exposure: A phased-based treatment for PTSD related to childhood abuse. *Journal of Consulting and Clinical Psychology, 70*, 1067–1074.

Cloitre, M., Stovall-McClough, C. K., Miranda, R., & Chemtob, C. M. (2004). Therapeutic alliance, negative mood regulation, and treatment outcome in child abuse-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 72*, 411–416.

Collier, P. (2006). African growth: Why a "big push"? *Journal of African Economies, 15*, 188–211.

Collier, P., Hoeffler, A., & Sambanis, N. (2005). The Collier–Hoeffler model of civil war onset and the case study project research design. In P. Collier & N. Sambanis (Eds.), *Understanding civil war: Vol. 1. Africa evidence and analysis* (pp. 1–33). Washington, DC: World Bank.

Davidson, L. M., & Baum, A. (1986). Chronic stress and post traumatic stress disorders. *Journal of Consulting and Clinical Psychology, 54*, 303–308.

Davidson, L. M., Baum, A., & Collins, D. L. (1982). Stress and control-related problems at Three Mile Island. *Journal of Applied Social Psychology, 54*, 303–308.

Demick, J., & Wapner, S. (1988). Children-in-environments: Physical, interpersonal, and sociocultural aspects. *Children's Environments Quarterly, 5*, 54–62.

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Reiser, M., Murphy, B. C., Holmgren, R., et al. (1997). The relations of regulation and emotionality to resiliency and competent social functioning in elementary school children. *Child Development, 68*, 367–383.

Eisenberg, N., Hofer, C., & Vaughan, J. (2007). Effortful control and its socioemotional consequences. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 287–306). New York: Guilford.

Eisenberg, N., & Morris, A. S. (2002). Children's emotion-related regulation. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 190–229). Amsterdam: Academic Press.

Eisenberg, N., Smith, C. L., Sadovsky, A., & Spinrad, T. L. (2004). Effortful control: Relations with emotion regulation, adjustment, and socialization in childhood. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 259–282). New York: Guilford.

Fabes, R. A. (1994). Physiological, emotional, and behavioral correlates of sex segregation. In C. Leaper (Ed.), *The development of gender and relationships* (pp. 19–34). San Francisco: Jossey-Bass.

Fleming, I., Baum, A., Gisriel, M. M., & Gatchel, R. J. (1987). Mediation of stress at Three Mile Island by social support. *Journal of Human Stress, 8*, 14–22.

Fritz, C. E., & Marks, E. S. (1954). The NORC studies of human behavior in disaster. *Journal of Social Issues, 10*, 26–41.

Gaffney, D. A. (2006). The aftermath of disaster: Children in crisis. *Journal of Clinical Psychology, 62*, 1001–1016.

Gerrity, E. T., & Steinglass, P. (2003). Relocation stress following catastrophic events. In R. J. Ursano, C. S. Fullerton, & A. E. Norwood (Eds.), *Terrorism and disaster: Individual and community mental health interventions* (pp. 259–286). Cambridge, UK: Cambridge University Press.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament* (pp. 249–272). New York: Plenum.

Goto, T., Wilson, J. P., Kahana, B., & Slane, S. (2006). The Miyake Island volcano disaster in Japan: Loss, uncertainty, and relocation as predictors of PTSD and depression. *Journal of Applied Social Psychology, 36*, 2001–2026.

Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology, 27*, 87–96.

Honwana, A. (2006). *Child soldiers in Africa*. Philadelphia: University of Pennsylvania Press.

Houlihan, D., Ries, B. J., Polusny, M. A., & Hanson, C. N. (2008). Predictors of behavior and level of life satisfaction of children and adolescents after a major tornado. *Journal of Psychological Trauma, 7*, 21–36.

Jones, L., & Kafetsios, K. (2005). Exposure to political violence and psychological well-being in Bosnian adolescents: A mixed method approach. *Clinical and Child Psychology and Psychiatry, 10*, 157–176.

Jones, R. T., Ribbe, D. P., & Cunningham, P. (1994). Psychosocial correlates of fire disaster among children and adolescents. *Journal of Traumatic Stress, 7*, 117–122.

Kamatsiko, V. V. (2007). *The future in our hands: Children displaced by conflicts in Africa's Great Lakes region*. Nairobi, Kenya: World Vision Africa.

Kimenyi, M. S., & Ndungu, N. S. (2005). Sporadic ethnic violence: Why has Kenya not experienced a full-blown civil war. In P. Collier & N. Sambanis (Eds.), *Understanding civil war: Vol. 1. Africa evidence and analysis* (pp. 123–156). Washington, DC: World Bank.

Klingman, A. (2002). Children under stress of war. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 359–380). Washington, DC: American Psychological Association.

Kocijan-Hercigonja, D., Rijavec, M., Marusic, A., & Hercigonja, V. (1998). Coping strategies of refuge, displaced, and non-displaced children in a war area. *Journal of Psychiatry, 52*, 45–50.

Kostelny, K. (2006). A culture-based, integrative approach: Helping war-affected children. In N. Boothby, A. Strang, & M. Wessells (Eds.), *A world turned upside down: Social ecological approaches to children in war zones* (pp. 19–38). Bloomfield, CT: Kumarian Press.

La Greca, A. M., & Prinstein, M. J. (2002). Hurricane and earthquakes. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 107–138). Washington, DC: American Psychological Association.

La Greca, A. M., Silverman, W. S., Vernberg, E. M., & Prinstein, M. J. (1996). Posttraumatic stress symptoms in children after Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology, 64*, 712–723.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (2002). *Helping children cope with disasters and terrorism*. Washington, DC: American Psychological Association.

La Greca, A. M., Silverman, W. K., & Wasserstein, S. B. (1998). Children's predisaster functioning as a predictor of posttraumatic stress following Hurricane Andrew. *Journal of Consulting and Clinical Psychology, 66*, 883–892.

Ladd, G. W., & Profilet, S. M. (1996). The Child Behavior Scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008–1024.

Lazarus, R. S., & Cohen, J. B. (1977). Environmental stress. In I. Altman & J. F. Wohlwill (Eds.), *Human behavior and the environment: Current theory and research* (Vol. 2, pp. 89–127). New York: Plenum Press.

Lengua, L. J., Long, A. C., Smith, K. I., & Meltzoff, A. N. (2005). Pre-attack symptomatology and temperament as predictors of children's responses to the September 11 terrorist attacks. *Journal of Child Psychology and Psychiatry, 46*, 631–645.

Levy-Shiff, R. (1982). The effects of father absence on young children in mother-headed families. *Child Development, 53*, 1400–1405.

Liddell, C. (1994). South African children in the year before school: Towards a predictive model of everyday behaviour. *International Journal of Psychology, 29*, 409–430.

Lifton, R. J., & Olsen, E. (1976). The human meaning of total disaster: The Buffalo Creek experience. *Psychiatry, 39*, 1–18.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543–562.

Machel, G. (1996). *Impact of armed conflict on children*. Retrieved November 8, 2008, from http://daccessdds.un.org/doc/UNDOC/GEN/N96/219/55/PDF/N9621955.pdf?OpenElement

Majtenyi, C. (2008). *Conflict in Kenya leaves thousands of homeless children in its wake*. Retrieved November 5, 2008, from http://www.voanews.com/english/archive/2008-04/2008-04-17-voa31.cfm

Makoloo, M. O. (2005). *Kenya: Minorities, indigenous peoples and ethnic diversity*. Retrieved March 7, 2009, from http://www.unhcr.org/refworld/docid/469cbfbe0.html

Masten, A. S. (1994). Resilience in individual development: Successful adaptation despite risk and adversity. In M. C. Wang & E. W. Gordon (Eds.), *Educational resilience in inner city America: Challenges and prospects* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Masten, A. S. (2004). Regulatory processes, risk and resilience in adolescent development. *Annals of the New York Academy of Sciences, 1021*, 310–319.

Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19*, 921–930.

Masten, A., Best, K., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*, 425–444.

Masten, A. S., & Obradović, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society, 13*, 9.

McDermott, B. M., & Palmer, L. J. (2002). Postdisaster emotional distress, depression, and event-related variables: Findings across child and adolescent developmental stage. *Australian and New Zealand Journal of Psychiatry, 36*, 754–761.

McFarlane, A. C. (1987). Posttraumatic phenomena in a longitudinal study of children following a natural disaster. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 764–769.

Miller, A. M., & Heldring, M. (2004). Mental health and primary care in a time of terrorism: Psychological impact of terrorist attacks. *Families, Systems, & Health, 22*, 7–30.

Moore, D. P., & Moore, J. W. (1996). Posthurricane burnout: An island township's experience. *Environment and Behavior, 28*, 134–155.

Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The role of the family context in the development of emotion regulation. *Social Development, 16*, 362–388.

Morris, A. S., Silk, J. S., Steinberg, L., Sessa, F. M., Avenevoli, S., & Essex, M. J. (2002). Temperamental vulnerability and negative parenting as interacting of child adjustment. *Journal of Marriage and Family, 64*, 461–471.

Morris, J., van Ommeren, M., Belfer, M., Saxena, S., & Saraceno, B. (2007). Children and the sphere standard on mental and social aspects of health. *Disasters, 31*, 71–90.

Moss, W. J., Ramakrishnan, M., Storms, D., Siegle, A. H., Weiss, W. M., Lejnev, I., et al. (2007). Child health in complex emergencies. *Bulletin of the World Health Organization, 84*, 58–64.

Ngige, L. W. (2008). *Impact of civil war on family and child survival: Perspectives on family resilience and challenges in Southern Sudan*. Presentation at Oklahoma State University.

Onyango, P. (1998). The impact of armed conflict on children. *Child Abuse Review, 7*, 219–229.

Peek, L. A., & Mileti, D. S. (2002). The history and future of disaster research. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 511–524). New York: Wiley.

Pfefferbaum, B., Pfefferbaum, R. L., Gurwitch, R. H., Nagumalli, S., Brandt, E. N., Robertson, M. J., et al. (2003). Children's response to terrorism: A critical review of literature. *Current Psychiatry Reports, 5*, 95–100.

Qouta, S., Punamäki, R. L., Miller, T., & El-Sarraj, E. (2008). Does war beget child aggression? Military violence, gender, age, and aggressive behavior in two Palestinian samples. *Aggressive Behavior, 34*, 231–244.

Rabalais, A. E., Ruggiero, K. J., & Scotti, J. R. (2002). Multicultural issues in the response of children to disasters. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 73–100). Washington, DC: American Psychological Association.

Ronen, T. (2002). Difficulties in assessing traumatic reactions in children. *Journal of Loss and Trauma, 7*, 87–106.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook on child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–166). New York: Wiley.

Rousseau, C., Drapeau, A., & Platt, R. (1999). Family trauma and its association with emotional and behavioral problems and social adjustment in adolescent Cambodian refugees. *Child Abuse and Neglect, 23*, 1263–1273.

Sagi-Schwartz, A., Seginer, R., & Abdeen, Z. (2008). Chronic exposure to catastrophic war experiences and political violence: Links to the well-being of children and families. *International Journal of Behavioral Development, 32*, 257–259.

Shannon, M. P., Lonigan, C. J., Finch, A. J., & Taylor, C. M. (1994). Children exposed to disaster I: Epidemiology of post-traumatic symptoms and symptom profiles. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 80–93.

Silverman, W. K., & La Greca, A. M. (2002). Children experiencing disasters: Definitions reactions, and predictors of outcomes. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 11–34). Washington, DC: American Psychological Association.

Smith, P., Perrin, S., Yule, W., Hacam, B., & Stuvland, R. (2002). War exposure among children from Bosnia-Herzegovina: Psychological adjustment in a community sample. *Journal of Traumatic Stress*, *15*, 147–156.

Sphere Project. (2004). *The sphere project: Humanitarian charter and minimum standards in disaster response* (2nd ed.). Washington, DC: Oxfam Publishing. Retrieved June 5, 2008, from http://www.sphereproject.org/component/option,com_docman/task,doc_view/gid,12/Itemid,203/lang,english/

Terranova, A. M., Boxer, P., & Morris, A. S. (2009). Factors influencing the course of posttraumatic stress following a natural disaster: Children's reactions to Hurricane Katrina. *Journal of Applied Developmental Psychology*, *30*, 344–355.

Thompson, M. P., Norris, F. H., & Hanacek, B. (1993). Age differences in the psychological consequences of Hurricane Hugo. *Psychology and Aging*, *8*, 606–616.

United Nations Children's Fund. (2002). *Facts for life: Disasters and emergencies*. Retrieved June 4, 2008, from http://www.unicef.org/ffl/pdf/factsforlife-en-part14.pdf

United Nations Children's Fund. (2008). *Hundreds of thousands of children suffer the consequences of the post-election crisis in Kenya*. Retrieved June 4, 2008, from http://www.unicef.org/media/media_42441.html

United Nations Human Settlements Program. (2003). *The challenge of slums; Global report on human settlements 2003*. London: Author.

Vernberg, E. M., La Greca, A. M., Silverman, W. K., & Prinstein, M. (1996). Predictors of children's post-disaster functioning following Hurricane Andrew. *Journal of Abnormal Psychology*, *105*, 237–248.

Wapner, S. (1987). A holistic, developmental, systems-oriented environmental psychology: Some beginnings. In D. Stokols & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 1433–1465). New York: Wiley.

Wapner, S., & Demick, J. (1998). Developmental analysis: A holistic, developmental, systems-oriented perspective. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 761–805). New York: Wiley.

Wapner, S., & Demick, J. (2002). The increasing contexts of context in the study of environment and behavior relations. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 3–14). New York: Wiley.

Wasserstein, S. B., & La Greca, A. M. (1998). Hurricane Andrew: Parent conflict as a moderator of children's adjustment. *Hispanic Journal of Behavioral Sciences*, *20*, 212–224.

Weems, C. F., Pina, A. A., Costa, N. M., Watts, S. E., Taylor, L. K., & Cannon, M. F. (2007). Predisaster trait anxiety and negative affect predict posttraumatic stress in youths after Hurricane Katrina. *Journal of Consulting and Clinical Psychology*, *75*, 154–159.

World Bank. (2007). *World development indicators*. Washington, DC: World Bank. Retrieved June 6, 2008, from http://siteresources.worldbank.org/DATASTATISTICS/Resources/table2_1.pdf

Yorbik, O., Akbiyik, D. I., Kirmizigul, P., & Sohmen, T. (2004). Post-traumatic stress disorder symptoms in children after the 1999 Marmara earthquake in Turkey. *International Journal of Mental Health*, *33*, 46–58.

Yule, W., & Williams, R. (1990). Post-traumatic stress reactions in children. *Journal of Traumatic Stress*, *3*, 279–295.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1129–1141

Impact of Maternal Posttraumatic Stress Disorder and Depression Following Exposure to the September 11 Attacks on Preschool Children's Behavior

Claude M. Chemtob, Yoko Nomura,
and Khushmand Rajendran
Mount Sinai School of Medicine

Rachel Yehuda
*Mount Sinai School of Medicine and
Bronx Department of Veterans
Affairs Medical Center*

Deena Schwartz
Mount Sinai School of Medicine

Robert Abramovitz
Jewish Board of Family and Children's Services

To evaluate whether conjoined maternal posttraumatic stress disorder (PTSD) and depression are associated with increased behavioral problems among terrorism-exposed preschool children (N = 116; 18–54 months), this study compared clinically significant child behavioral problem rates among the preschool children of mothers with PTSD and depression, depression alone, and neither disorder. Behavioral problems were independently rated by mothers and preschool teachers. Maternal depression and PTSD, relative to maternal depression alone, and to neither disorder, were associated with substantially increased child problems. Notably, maternal depression and PTSD were associated with increased emotional reactivity (relative risk [RR] = 5.9 by mother's and 3.4 by teacher's reports) and aggressive behavior problems (RR = 11.0 by mother's and RR = 5.9 by teacher's reports). This was corroborated by teacher ratings. Implications for intervening with terrorism-exposed preschool children are discussed.

Although preschool children are highly vulnerable to the effects of traumatic events, the effects of maternal psychopathology on preschool children exposed to terrorism are not as yet well understood (Chemtob, Nomura, & Abramovitz, 2008; DeVoe, Bannon, & Klein, 2006; Fremont, 2004; Laor, Wolmer, & Cohen, 2001; Pfefferbaum, 2005; Pine, Costello, & Masten, 2005; Thabet, Karim, & Vostanis, 2006).

Consistent with McEwen's (1998) allostatic load hypothesis, preschool children with conjoined exposure to World Trade Center (WTC) attack-related events and to other traumatic events have been found to be far more likely to develop clinically significant behavioral problems than children with WTC-attack exposure alone, or with other trauma exposure alone, compared to children without either type of trauma exposure (Chemtob et al., 2008). Clearly, exposure to the WTC attacks is likely to have also impacted the mothers of these children and may have caused some of these mothers to develop psychological disorders.

It is likely that very young children are affected by terrorism both as a result of the effects of direct exposure and, more indirectly, as a result of the impact of terrorism on their mothers' psychological functioning. However, little is known about whether maternal psychopathology associated with exposure to terrorism may be associated with increased rates of child behavioral problems.

Limitations of Current Research

The extant studies of the effects of maternal psychopathology on trauma-exposed children have tended to focus on older school-age children (cf. McCloskey & Walker, 2000; Najarian, Goenjian, Pelcovitz, Mandel, & Najarian, 1996; Smith, Perrin, Yule, & Rabe-Hesketh, 2001; Vila et al., 2001). Also, studies have usually confounded the effects of maternal and child event exposure (cf. Green et al.,

---

Correspondence concerning this article should be addressed to Claude M. Chemtob, One Gustave L Levy Place, Box 1230, New York, NY 10029. Electronic mail may be sent to claude.chemtob@mssm.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0008

1991; Kilic, Ozguven, & Sayil, 2003; Wolmer, Laor, Gershon, Mayes, & Cohen, 2000). Moreover, research on the effects of disaster-related exposure on children has generally not distinguished the effects of disaster exposure from the effects of exposure to other lifetime trauma (cf. Chemtob et al., 2008), nor, to the best of our knowledge, have any of these studies specifically examined the effects of co-occurring maternal disorders on disaster-related child psychopathology.

In addition, nearly all the above studies relied on mothers as the sole informants about their preschool children. This reliance on mothers introduces a potential confound because of the possibility that maternal psychopathology would create systematic biases in the ratings of children's behavioral problems (Breslau, Davis, & Prabucki, 1988; Fergusson, Lynskey, & Horwood, 1993; Renouf & Kovacs, 1994). These maternal ratings of children may reflect the influence of their depression or posttraumatic stress disorder (PTSD) symptoms to an unknown degree and may result in rating biases. Teacher ratings are not necessarily more veridical than those of mothers, because teachers interact with children in a different context with different demands (Richters & Pellegrini, 1989). Nevertheless, examining both maternal and teacher ratings of child behavior would provide an opportunity to evaluate independent ratings of the child's functioning.

Finally, studies of the effects of maternal psychopathology on children have largely focused on maternal depression (Galea et al., 2002). To the best of our knowledge, no studies have directly contrasted the effects of maternal depression alone with the effects of co-occurring maternal depression and PTSD on terrorism-exposed very young children. Comparing the effects of maternal depression alone with the effects of conjoined depression and PTSD is potentially conceptually important. PTSD is a disorder in which the dysregulation of the response to threat is highly salient (Chemtob, Roitblat, Hamada, Carlson, & Twentyman, 1988; Yehuda, 2002). Chemtob et al. (1988) described the salience of threat detection as an attentional priming effect. The priming hypothesis has subsequently been empirically supported (Michael & Ehlers, 2007; Michael, Ehlers, & Halligan, 2005; Vythilingam et al., 2007). In addition, notably, there is evidence that women may have a lower threshold to threat appraisal compared to men (Olff, Langeland, Draijer, & Gersons, 2007). Thus, it can be expected that as a result of direct exposure to a direct threat to self and to child, a mother with PTSD will subsequently be more vigilant with respect to the presence of a possible threat and to have a lower threshold for identifying events as dangerous to herself, and to her child's safety.

Thus, a mother's PTSD can be expected to result in an increased emotional regulation burden for both mother and very young child (Hirshfeld, Biederman, Brody, Faraone, & Rosebaum, 1997) as the mother seeks to restore a sense of safety following exposure to terrorism. In short, a mother with PTSD may face a greater challenge in creating a sense of safety in her child because of her sensitivity—subtle or more overt—to potential danger. As a result, she may cue her young child to the presence of danger even when it is not present.

Depressed mothers may exhibit a lack of control over their environment, an inability to draw boundaries and inconsistency in disciplining their children (Kochanska, Kuczynski, Radke-Yarrow, & Welsh, 1987; Zahn-Waxler, Iannotti, Cummings, & Denham, 1990). They may also show more withdrawn or uneven affect and may have trouble interacting appropriately with infants and toddlers (Cohn & Tronick, 1989; Downey & Coyne, 1990; Murray, 1992; Stein et al., 1991). Thus, maternal depression is likely to be expressed through reduced maternal engagement with the child and reduced availability to the child, which may result in a lack of support for the child's emotional and behavioral regulation. Indeed, it has been suggested that infants who receive inconsistent responses from their caregivers may not develop the ego strength and ability to recover from the effect of negative interactions (Osofsky & Eberhart-Wright, 1988). Similarly, a depressed mother may be less attuned to the child's needs for contact, support, and encouragement. This suggests that the lack of availability of a depressed mother may restrict a child's ability to restore emotional and behavioral regulation following exposure to the threat of terrorism.

Thus, it could be argued that combined depression and PTSD might be expected to have a greater impact on a mother's ability to support her child's emotional and behavioral regulation. A mother's relative lack of availability (related to depression) to support her child's recovery, and her disposition to feeling unsafe and to overidentify potential threats (related to her PTSD) would compound the difficulty for a child of restoring his or her sense of safety after exposure to terrorism. This would then be predicted to result in increased problems in emotional and behavioral regulation, leading to a higher rate, and larger number, of child behavioral problems being associated with conjoined maternal PTSD and depression. Conversely, maternal depression *alone* may be sufficiently pervasive in its effects on child well-being in young children, because of its effects on attachment and child security, that the additional burden of maternal PTSD would have no added consequence for the child's ability to regulate emotion and behavior. In this instance, it would be expected that there would be no difference in rate, or number, of behavioral problems between the children of mothers with both PTSD and depression and the children of mothers with only depression relative to the children of mothers with neither disorder.

The Current Study

Accordingly, this study compared the impact of maternal depression alone to the effects of maternal depression and PTSD together on child behavior problems. We examined data from our study of the long-term impact of direct exposure to the WTC attacks on clinically significant behavioral problems in preschool children (Chemtob et al., 2008). In keeping with our interest in examining the relative effects of disaster exposure and lifetime exposure to other types of trauma, we had assessed maternal exposure to both the WTC attacks and to other trauma. We had also assessed maternal depression and PTSD. Finally, because a significant limitation of prior research on the impact of maternal psychopathology on very young children has been its reliance on single informants (usually mothers), we had obtained independent child behavior ratings from the preschool teachers of the children we studied. The teacher ratings provided the opportunity to obtain evidence from independent observers, who were not aware of the mother's mental state, about the children's behavior problems.

To address our primary question, we divided mothers into three groups: (a) mothers with neither disorder, (b) mothers with depression in the absence of PTSD, and (c) mothers with both depression and PTSD. We hypothesized that after removing the effects of child and mother exposure to the WTC attacks, and to other trauma, as well as demographic confounders, co-occurring maternal depression and PTSD would be associated with higher rates of clinically significant child behavioral problems compared to children of mothers with depression only, and compared to children of mothers with neither PTSD nor depression. We also hypothesized that the children of mothers with both depression and PTSD would have the highest number of clinically significant behavior problems. We focused on clinically significant behavior problems because they indicate child problems severe enough to be of possible public health significance.

Method

Participants

This study was reviewed and approved by the Mount Sinai School of Medicine's Institutional Review Board. It was conducted between March 2003 and December 2005 (a mean of 35 months after the WTC attacks) to assess the longer term impact of WTC-attack exposure on preschool children age 5 years or younger at the time of the WTC attacks. Families with children born between September 11, 1996, and September 11, 2002, were included if: (a) they lived in Lower Manhattan or (b) attended preschool or day care in Lower Manhattan on September 11, 2001 (the day of the WTC attacks). Families were recruited using extensive outreach in the Lower Manhattan area, primarily to preschools. (See Figure 1 for the geographical location of preschools relative to the site of the WTC attacks.)

After consent was obtained, mothers completed questionnaires about their child and themselves that included demographic information, exposure to the WTC attacks, exposure to other traumatic events, and emotional and behavioral functioning (see the Measures section). Children and mothers were assessed, and if symptomatic, they were offered free diagnostic and counseling services. The child's preschool teacher, who was blind to maternal report of the child behavioral problems and maternal psychopathology, also completed a questionnaire about the child's behavior. Teacher and parent questionnaires were administered independently from each other.

A total of 116 children and their mothers were included in the study (Chemtob et al., 2008). Three children of mothers who had PTSD only were excluded from these analyses, leaving 113 children. Participants who dropped out of the study did not differ on age, gender, race, or socioeconomic status (SES) from completers.

Measures

Demographics. Age, ethnicity of the child, and maternal education in years were reported by the mother. Children's age and time elapsed since the

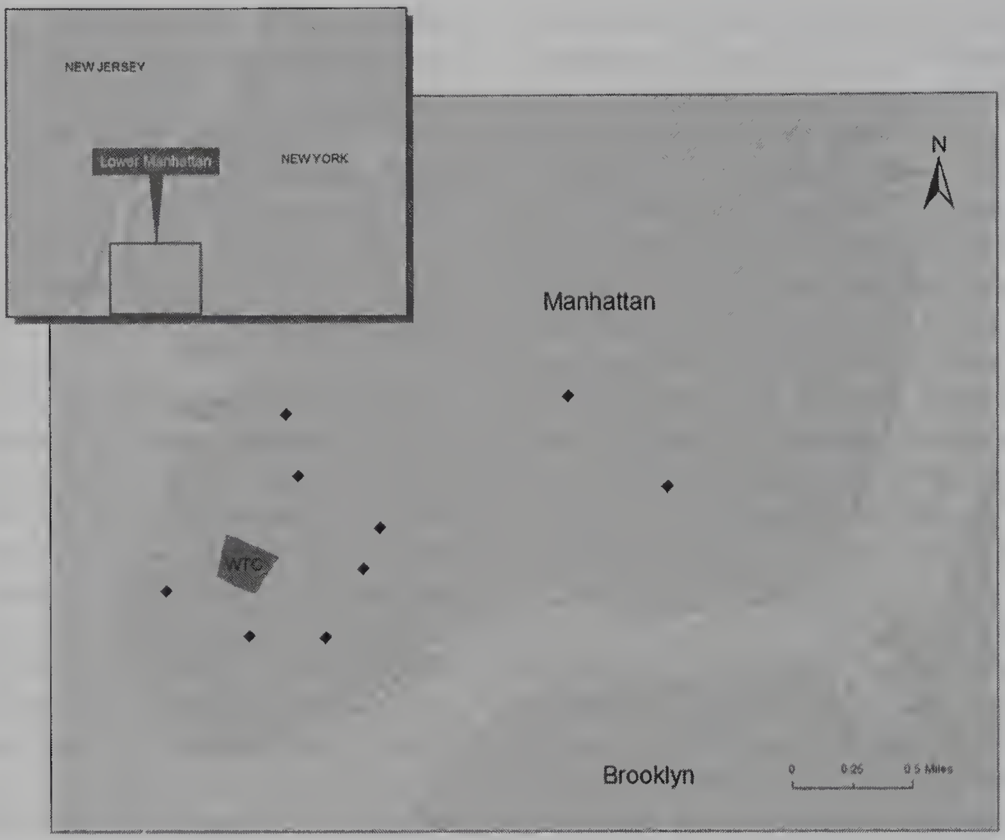


Figure 1. The geographical location of preschools relative to the World Trade Center (WTC).

WTC attacks were calculated based on the date of assessment, the child's date of birth, and September 11, 2001, respectively. SES was measured by maternal education, a reliable and valid index of SES (Bollen, Glanville, & Stecklov, 2007; Liberatos, Link, & Kelsey, 1988).

Maternal exposure to the WTC attacks and to other trauma. Maternal WTC-attack exposure was measured using the mother's dichotomous responses to six questions about direct exposure to high-intensity WTC-attack-related events. Respondents were asked whether they personally saw a plane hit a tower, saw the tower collapse, saw injured people, saw dead bodies, saw people jumping out of the building, and were caught in the debris or smoke. Responses to the six questions were summed to denote level of exposure. We used the trauma event checklist of the Posttraumatic Stress Diagnostic Scale (PDS) to measure lifetime history of other trauma exposure (Foa, 1995).

Child exposure to the WTC attacks and to other trauma. Mothers reported on the extent to which their children had been directly exposed to the WTC attacks. We created an index of child exposure to the same six high-intensity attack-related events. Child lifetime history of exposure to trauma other than the WTC attacks was measured using a modified version of the Traumatic Events Screening Inventory (TESI; Ghosh Ippen et al., 2002). The instrument measures exposure to traumatic events such as natural disasters, interpersonal losses, serious accidents, severe illnesses or injuries, animal attacks, exposure to war or terrorist acts, and exposure to suicide or attempted suicide.

Child behavior problems. Mothers used the preschool version of the Child Behavioral Checklist (CBCL/1.5–5; Achenbach & Rescorla, 2000), a widely used behavioral checklist with very good reliability and validity (Calkins & Dedmon, 2000; Gross et al., 2006; Mesman, Bongers, & Koot, 2001), to rate child behavior problems. In the normative study, test–retest reliability of syndrome scales of CBCL/1.5–5 ranged from .68 to .92 and it demonstrated good validity by its association with referral status of the child (Achenbach & Rescorla, 2000). Age and gender standardized *T* scores were calculated for the seven behavioral problem clusters (emotionally reactive, anxious/depressed, somatic complaints, withdrawn behavior, sleep problems, attention problems, and aggressive behavior). There were 5 children above the age of 5 at the time of assessment, 4 were 6 years old, and 1 was age 7. We followed the recommendation of the scale developer (Thomas Achenbach, personal communication, 2004) to use the preschool version of the scale for all the children in our sample. The preschool scale largely differs from the older child scale (CBCL 6–18; Achenbach & Rescorla, 2001) in that it omits a number of items particularly appropriate to adolescent children.

Preschool teachers, who were "blind" to maternal psychopathology and to the child behavioral ratings provided by mothers, completed the preschool version of the Teacher Report Form (TRF; Achenbach & Rescorla, 2000). The TRF behavior-rating checklist (Achenbach & Rescorla, 2000) parallels the CBCL and was designed for completion by teachers. The TRF has high test–retest reliability (mean $r = .81$) and good correlations with the CBCL (mean $r = .40$). The CBCL sleep problems cluster is not included in the teacher rating form of the CBCL.

Nevertheless, because sleep problems are associated with trauma exposure in children, and because they are of clinical importance, we present the parental ratings of sleep problems in Table 2 (see p. 1134). However, as this article is focused on using parent and teacher ratings, and because teachers do not rate sleep problems, the sleep problems cluster is not included in further analyses.

This article focuses on clinically significant preschool child behavioral problems. Scores ≥ 65 (93rd percentile) for the behavioral problem scales are considered clinically significant levels of problems for both the CBCL and the TRF. This cutoff was used to define clinically significant levels of behavior problems for both mother ratings and teacher ratings.

Maternal depression and PTSD. The Center for Epidemiologic Studies Depression Scale (CES–D: Radloff, 1977) was used to assess maternal depressive symptomatology. Probable maternal depression was dichotomized using the CES–D clinical cutoff score of 16, which is widely used (Breslau, 1985) for that purpose. The PDS (Foa, 1995) was used to measure maternal posttraumatic stress symptoms. The structure and content of the PDS parallel the *DSM–IV* PTSD diagnostic criteria (Foa, 1995). The PDS has a test–retest reliability kappa of .74 for PTSD diagnosis and convergent validity with a kappa of .65 with the Structured Clinical Interview (SCID) for *DSM–III–R*, with 82% agreement between the two measures (Foa, 1995). The sensitivity and specificity of the PDS with respect to SCID diagnosis is .89 and .75, respectively (Foa, 1995). Mothers were considered to have PTSD if they met all six diagnostic criteria—exposure to a traumatic event, persistent reexperiencing symptoms, persistent avoidance of the stimuli associated with the trauma, persistent arousal, duration of the disturbance for longer than 1 month, and clinically significant functional impairment.

Statistical Analysis

The major analyses are based on grouping the mothers according to whether they had: (a) neither depression nor PTSD, (b) depression alone, or (c) both depression and PTSD. Based on this subdivision, we first compared relevant demographic and trauma-exposure characteristics to detect potential confounding variables. Analyses were performed first based on mother's report. Confirmatory analyses were then conducted based on the behavioral ratings of teachers. Mother and teacher reports of child behavior were entered as dependent variables in separate models.

To examine the rates of clinically significant child behavior problems by maternal psychopathology, univariate analyses were conducted using chi-square tests for overall group difference followed by chi-square linear-by-linear association, also known as the Mantel Haenszel chi-square test (Dosemeci & Benichou, 1998; Hakulinen, 1981) for overall trend to evaluate a dose–response relation between maternal psychopathology groups (neither, depression only, and both depression and PTSD) and each child behavioral problem cluster.

Next, to examine the risk for each behavioral problem in children of mothers with PTSD and depression and in children of mothers with depression alone, hierarchical logistic regression was performed after creating two dummy variables for maternal psychopathology: (a) for mothers with depression only and (b) for mothers with both depression and PTSD. The group of mothers with neither PTSD nor depression served as the reference group. The effects of potential confounders described earlier were first included before the two dummy variables for maternal psychopathology group were entered in the model. The risk of having each behavioral problem (emotionally reactive, anxious/depressed, somatic complaints, withdrawn behaviors, aggressive behaviors, and attention problems), as measured by odds ratio, for children of mothers with only depression and children of mothers with comorbid depression and PTSD, as compared to the reference group, based on mother's and teacher's report, was first obtained. As odds ratio is only an approximation of the relative risk and could be inflated due to high occurrence of the problems, we converted our odds ratio into relative risk. We have used the converting formula presented by Zhang & Yu (1998):

$$RR = \frac{OR}{(1 - P_0) + (P_0 \times OR)},$$

where P_0 indicates the incidence of the outcome of interest in the nonexposed group (i.e., children of mothers with neither depression nor PTSD). Ninety-five percent confidence intervals (CI) were also converted, accordingly.

As children were drawn from 10 different preschools/day-care centers, and some were from the same family (*n* = 13), the assumption of independent observations underlying standard errors and CIs may be violated. To account for data clustering, standard errors for rates of behavioral problems were computed using the Taylor series linearization method, while standard errors and CIs for parameters estimated from the logistic analyses were computed by using the method of weighted maximum likelihood, adapted for survey and clustered data through implicit Taylor series (Binder, 1992; Lin & Wey, 1989). Missing data were handled through listwise deletion. There were no missing data on the key study variables and minimal missing data in the control variables: child exposure to trauma other than 9/11 related (*n* = 2, 1.7%) and maternal education (*n* = 8, 6.9%).

Finally, as shown in Figure 3 (see p. 1136), we counted the number of clinically significant behavioral problems rated by mothers and by teachers for each child. One-way analyses of covariance (ANCOVAs; with child age, maternal education, and time passed since the WTC attacks used as covariates) were conducted separately for mother and teacher ratings. We normalized the outcome variables (number of behavioral problems based on mothers and teachers reports) by logarithm transformation to satisfy the assumptions of normal distribution and heteroscedasticity, that is, a condition where the variances of dependent variable are not equal for every value of predictor or grouping variable (Kotler & Brown, 1990; Schmidt, 2000). Because we used logarithm-transformed values in our analysis to remove heteroskedasticity, the numbers reported were transformed back in the original integer scale. We provide the *F* statistics from the ANCOVA based on both the transformed and nontransformed variable. We present the diagram (Figure 3) that shows the mean number of clinically significant behavioral problems among different maternal psychopathology groups based on nontransformed outcome values, as they were very similar and some readers may find the nontransformed means more clinically meaningful.

Results

Table 1 presents participant characteristics by maternal psychopathology category. There were no differences in age, gender distribution, maternal education, or exposure to the WTC attacks. However, mothers with depression and PTSD, and their children, were significantly more likely to have experienced other traumatic events. Approximately, 12% of the children in the sample met criteria for clinically significant emotionally reactive behaviors and for somatic complaints, 10% for anxious/depressed, 9% for withdrawn behavior and for attention problem, and 8% for aggressive behavior clusters.

Table 1
Characteristics of Participants by Maternal Psychopathology Group

	Maternal psychopathology group		
	Neither (*n* = 75), *M* (*SE*)	Depression (*n* = 24), *M* (*SE*)	Comorbid (*n* = 14), *M* (*SE*)
Child age	3.95 (1.10)	3.79 (0.93)	3.50 (0.65)
Child gender (girls), % (*n*)	48.00 (36)	54.17 (13)	64.29 (9)
Maternal education[a]	13.65 (3.01)	14.13 (3.81)	12.64 (2.81)
Maternal exposure to the WTC-related trauma[b]	0.27 (0.55)	0.50 (0.83)	0.50 (0.85)
Maternal exposure to other trauma[b]	1.27 (1.62)	0.87 (0.97)	3.42 (1.55)[c]
Child exposure to the WTC-related trauma[b]	0.44 (0.85)	0.83 (1.23)	0.43 (0.65)
Child exposure to other trauma[b]	0.92 (1.24)	1.21 (1.41)	2.93 (1.94)[d]
Time passed since WTC attacks (months)	35.31 (11.93)	32.72 (10.60)	36.69 (9.24)

Note. *n* value may slightly vary due to missing values. WTC = World Trade Center.
[a]Number of years of education. [b]Number of different exposure. [c]Significant group difference based on analysis of variance (ANOVA), $F(2, 110) = 6.1$, $p < .001$. [d]Significant group difference based on ANOVA, $F(2, 110) = 8.7$, $p < .001$.

Table 2

Rates per 100 (SE) of Mothers' Reports of Behavioral Problems in Children According to Whether the Mother Has Neither, Depression Only, or Both Depression and Posttraumatic Stress Disorder (PTSD)

	Maternal psychopathology								
	Group 1	Group 2	Group 3	Overall difference			Linear-by-linear associations[b]		
Child behavior problems[a]	Neither (n = 75) % (*SE*)	Depression only (n = 24) % (*SE*)	Depression & PTSD (n = 14) % (*SE*)	χ^2	*df*	*p* value	χ^2	*df*	*p* value
Emotionally reactive	5.33 (0.03)	8.70 (0.06)	57.14 (0.14)	29.34	2	.0001	22.47	1	.00001
Anxious/depressed	4.00 (0.02)	13.00 (0.07)	35.71 (0.13)	13.74	2	.001	13.38	1	.0002
Somatic complaints	6.66 (0.03)	22.70 (0.08)	28.57 (0.12)	7.68	2	.02	7.13	1	.008
Withdrawn behavior	6.66 (0.03)	8.70 (0.06)	21.43 (0.11)	3.16	2	.21	3.12	1	.08
Sleep problems	4.00 (0.02)	29.00 (0.10)	29.00 (0.13)	14.46	2	.001	11.93	1	.001
Aggressive behavior	2.66 (0.02)	4.54 (0.04)	42.85 (0.14)	26.04	2	.001	19.05	1	.00001
Attention problems	5.33 (0.03)	13.64 (0.07)	21.43 (0.11)	4.45	2	.11	4.32	1	.04

Note. The n values may vary due to missing values.

[a] A score of 65 or more on each Child Behavioral Checklist cluster was used to indicate behavioral problem symptoms. Depression is defined as scores above 16 on the Center for Epidemiologic Studies Depression Scale; PTSD is defined as positive on all six symptom clusters indicated by the PDS. [b] Chi-square linear-by-linear association is used to evaluate whether rate of behavior problems increases in a dose–response fashion.

Table 3

Rates per 100 (SE) of Teacher's Report of Her Child's Behavioral Problems in Children According to Whether the Mother Has Neither, Depression Only, or Both Depression and Posttraumatic Stress Disorder (PTSD)

	Maternal psychopathology								
	Group 1	Group 2	Group 3	Overall difference[b]			Linear-by-linear associations[c]		
Child behavior problems[a]	Neither (n = 64) % (*SE*)	Depression only (n = 24) % (*SE*)	Depression & PTSD (n = 14) % (*SE*)	χ^2	*df*	*p* value	χ^2	*df*	*p* value
Emotionally reactive	14.29 (0.04)	18.18 (0.06)	42.86 (0.14)	6.05	2	.005	4.96	1	.026
Anxious/depressed	14.29 (0.04)	18.18 (0.07)	35.71 (0.13)	3.54	2	.014	13.39	1	.08
Somatic complaints	6.35 (0.03)	9.10 (0.08)	14.29 (0.10)	1.01	2	.07	7.13	1	.32
Withdrawn behavior	9.52 (0.04)	13.64 (0.06)	28.57 (0.13)	3.65	2	.017	3.12	1	.07
Aggressive behavior	7.94 (0.03)	22.73 (0.04)	57.14 (0.14)	19.04	2	.00001	19.05	1	.00001
Attention problems	7.94 (0.03)	22.73 (0.07)	28.57 (0.13)	5.70	2	.04	4.32	1	.02

[a] A score of 65 or more on each Child Behavioral Checklist cluster was used to indicate behavioral problem symptoms. Depression is defined as scores above 16 on the Center for Epidemiologic Studies Depression Scale; PTSD is defined as positive on all six symptom clusters indicated by the PDS. [b] Fisher's exact test was applied for chi-square statistics when at least one of the cells had ≤ 5 observations. [c] Chi-square linear-by-linear association is used to evaluate whether rate of behavior problems increases in a dose–response fashion.

Table 2 presents the rate of each behavioral problem cluster in children, based on mother's report. When the rates of behavioral problems in children associated with each of three groups of mothers were compared, there were significant differences for all child behavioral problem clusters, except for withdrawn behavior and attention problems. There was a significant dose–response relation (i.e., where dose is defined as the number of maternal diagnoses) for all child behavioral domains, except withdrawn behavior. When the analyses were repeated using teacher reports, we found that the rates of child behavior problems associated with the maternal psychopathology groups were similar to those found based on the mother ratings of the children. However, teachers reported observing fewer child somatic complaints and more aggressive behavior. For teachers, as seen in Table 3, a significant dose–response relation in emotionally reactive, aggressive behavior, and attention problems was found. Fisher's exact test was applied in all analyses.

To confirm the increased risk based on maternal psychopathology category, a multiple logistic regression was performed. To minimize bias due to multicollinearity between mothers' and children's exposure to other trauma and maternal psychopathology group, first we entered potential confounders, including a priori determined demographic confounders and mothers' and children's exposure to the WTC and other trauma and then entered maternal psychopathology dummy variables in the model. Figure 2 presents the relative risks (RR) based on the mothers' ratings and the teachers' ratings of child behavior side by side. Whiskers at the end of the bars represent standard errors. Compared to the reference group of children of mothers with neither depression nor PTSD, depression and PTSD occurring together significantly increased the risk for aggressive behavior (RR = 11.0), emotionally reactive (RR = 5.9), anxious/depressed (RR = 4.8), and somatic complaints (RR = 4.1). In contrast, there was no significantly increased risk for behavioral problems in children of mothers with only depression with the exception of somatic complaints (RR = 3.4).

Similar patterns were observed across rater type (mother, teacher) in emotionally reactive behavior and aggressive behavior. However, teachers were more likely to observe problems in withdrawn behavior (RR = 9.0, p = .006) among the children of mothers with depression and PTSD. Based on mothers reports, significant increased risk of somatic complaints were reported in children whose mothers had both PTSD and depression (RR = 4.1, p = .03) and children whose mothers had only depression (RR = 3.4, p = .04), whereas no significant increased risk was reported based on teachers report.

Finally, as shown in Figure 3, we present the mean number of clinically significant behavioral problems (i.e., nontransformed), as rated by mothers and by teachers for each child. One-way ANCOVA (with child age, maternal education, and time passed since the WTC attacks used as covariates) shows that there was a significant difference in number of child behavioral problems by mater-

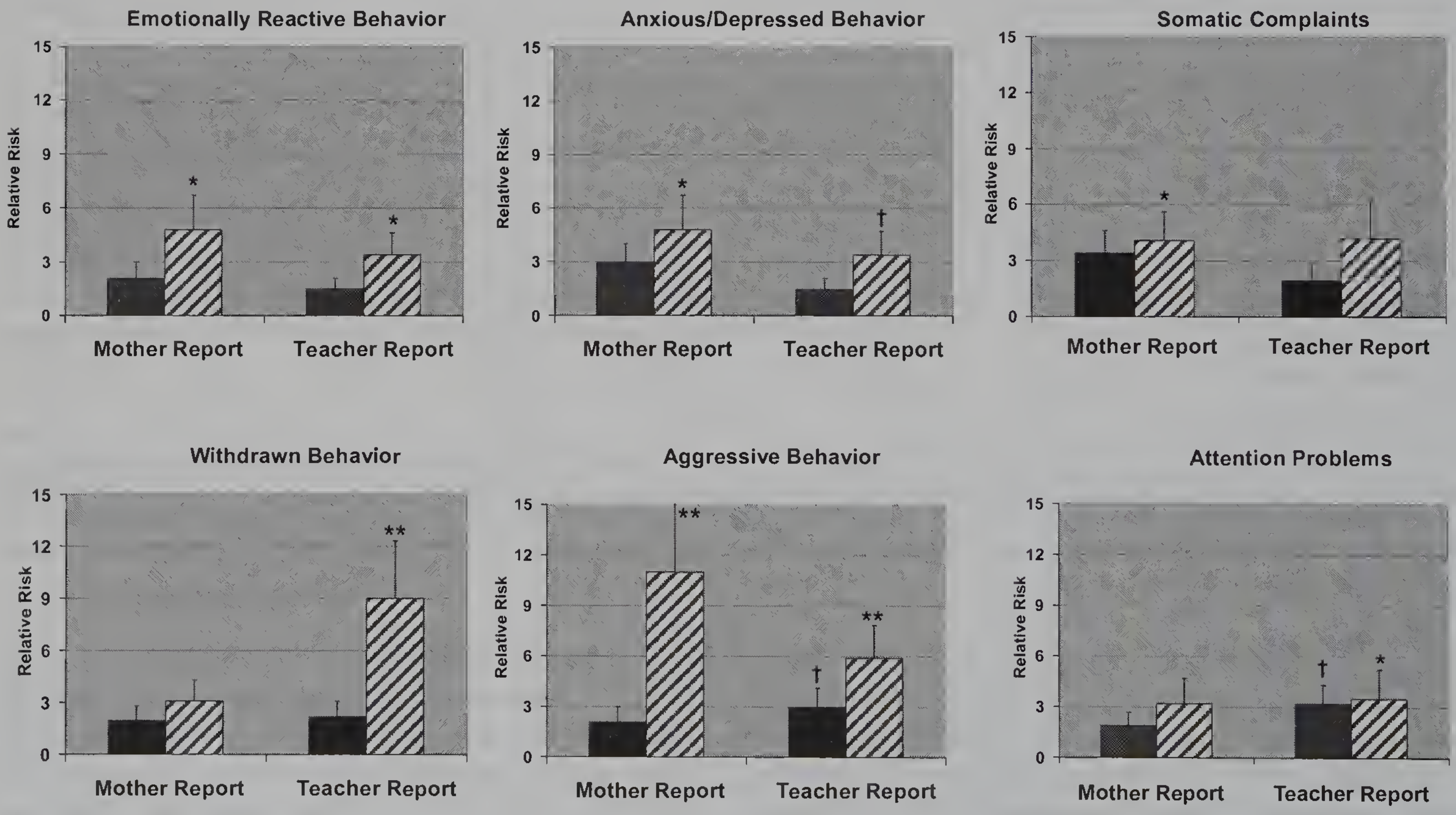


Figure 2. Risk for behavioral problems in children of mothers with only depression and with both depression and posttraumatic stress disorder (PTSD), based on mother's and teacher's reports.

Note. Striped bars represent both PTSD and depression. Bars with block print represent depression only. Risk in two groups (depression only and PTSD and depression) was calculated relative to that in children whose mother had neither depression nor PTSD. The effects of exposure to the World Trade Center and other trauma as well as potential confounder effect were statistically adjusted for in multiple logistic regression model.

$^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

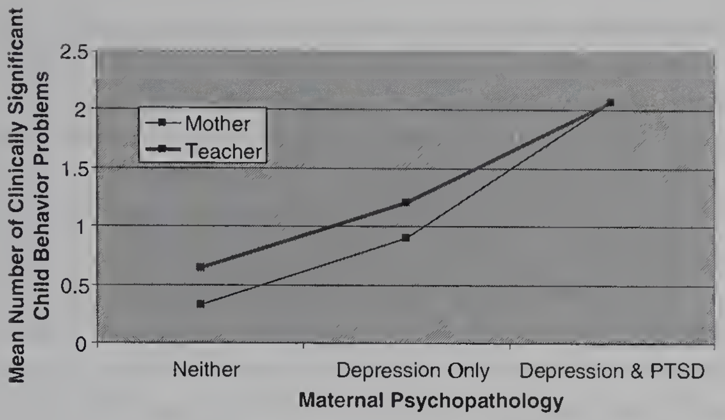


Figure 3. Mean number of clinically significant child behavior problem clusters (as rated by mothers and teachers) by maternal psychopathology group.

nal psychopathology category, and by mother's report based on nontransformed outcome variable, $F(2, 110) = 15.1$, $p = .001$, and transformed outcome variable, $F(2, 110) = 4.1$, $p = .002$. There was a similar significant difference in number of child behavioral problems by maternal psychopathology category by teacher's report based on the nontransformed outcome variable, $F(2, 96) = 5.4$, $p = .006$, and the transformed outcome variable, $F(2, 96) = 3.7$, $p = .004$. Post hoc testing, based on the transformed outcome variable, demonstrated that children of mothers with depression alone had a larger number of behavioral problem clusters that met clinical significance than children of mothers with neither PTSD nor depression, rated by the mother ($p = .04$). Children of mothers with both depression and PTSD had a significantly larger number of behavioral problem clusters that met clinical significance criteria than children of mothers with depression alone, rated by mother ($p = .04$) or by teacher ($p = .03$). A test of the difference between the regression slopes associated with each informant type showed no significant difference ($p = .78$).

Discussion

This study has three main findings. First, the children of mothers with co-occurring probable depression and PTSD had the highest rate of clinically significant behavior problems compared to children of mothers with depression and to the children of mothers with neither disorder. Second, the linear-by-linear association analysis showed a significant dose–response relation in emotionally reactive, aggressive behavior, and attention problems in their children, based on both mother's and independent teacher's reports. Mothers' reports additionally showed a dose–response relation in anxious/depressed and somatic complaints in their children. Third, similarly controlling for child and mother trauma exposure, children of mothers who had depression and PTSD were found to meet the cutoff for clinical significance for a larger number of behavior problem clusters, as rated by both mothers and teachers. Maternal and teacher reports were strikingly concordant in this regard highlighting that the differences in rates of child behavioral problems by maternal psychopathology grouping were not due to rating biases associated with maternal psychopathology.

Emotionally reactive behavior problems and aggressive behavior problems appear to be the most saliently associated with co-occurring maternal PTSD and depression. Notably, they are also the behavioral problems on which mothers and teachers have the highest observed concordance. The emotionally reactive behavior cluster indexes emotional regulation (Achenbach & Rescorla, 2000), whereas the aggressive behavior cluster indexes the behavioral regulation of outer-directed negative behaviors (cf. Barkley et al., 2000; Calkins & Dedmon, 2000; Shaw, Owens, Giovannelli, & Winslow, 2001). These two behavior clusters are of high public health significance. The results suggest that co-occurring maternal depression and PTSD interfere substantially more with the mother's ability to assist her young child in accomplishing the tasks associated with emotional and behavioral regulation than maternal depression alone.

Our findings may be interpreted either as consistent with the proposition that having two diagnoses is generally more impairing for mothers than having one, or to the additive but different effects of maternal PTSD and depression. The question of differential effects of maternal "dose" or "type" of psychopathology should be pursued in future studies. Little is known about whether a mother having co-occurring PTSD and depression results in greater generalized maternal impairment and reflects a nonspecific additive burden (a greater dose) or alternately whether maternal PTSD and depression each interfere with specific maternal functions and result in greater impact on the child because multiple maternal functions are affected (type). To illustrate, a depressed mother may be selectively unable to praise positive behavior in her child. A mother with PTSD may selectively respond to potential threat in her child's play and increase the child's anxiety. A mother with depression and PTSD may be so affected by her own distress that she is generally unavailable to her child, reflecting a generalized effects of dose (number of psychopathologic problems), or she may be both unable to praise positive behavior, reducing her support for child's positive coping and selectively overrespond to perceived threats increasing her child's anxiety. This latter possibility would suggest specificity (type) effects related to maternal PTSD and depression. The question is further complicated because children may react with specific behavior problems in association with specific maternal problems. However, it is also possible that they would have more problems in general as maternal number of problems increases without specific expression, reflecting whether the mother has depression, PTSD, or both.

These are important questions for further research with traumatized very young children and their mothers to explore. Such studies would be materially advanced by incorporating observational studies of dyadic interaction. Indeed, if there are different patterns of dyadic interaction associated with conjoined PTSD and depression, compared to the other two groups, and these dyadic interaction differences are associated with distinct child behavioral problems, this would suggest that specific behavioral problems in offspring may have different etiologies and modes of transmission. Similarly, it would be useful to examine biological markers associated with the intergenerational transmission of traumatic stress vulnerability (Yehuda & Bierer, 2008). Although they are highly informative, the data on which this study is based do not allow us to address the question of how maternal depression and PTSD specifically exert their effects on preschool children, and this remains an important research frontier.

Study Limitations

Although cross-sectional in design, our study included consideration of the effects of maternal and child lifetime history of other trauma exposure. This aspect of trauma exposure has not previously been taken into account systematically to the best of our knowledge. Our findings are strongly buttressed by the concordance of independent ratings of child problems by parent and teacher. Future studies should further extend these descriptive findings by including observational and behavioral measures to characterize differences among children and mothers in more specific detail. Also, future studies should specifically examine possible differences in dyadic interaction between child and mother that might be associated with maternal psychopathology. Such observational and behavioral studies may help elucidate mechanisms that distinguish the effects of conjoined depression and PTSD from those of maternal depression alone. We assessed maternal PTSD and depression using well-validated self-report measures. Nevertheless, future studies should also include SCIDs that assess the full range of maternal psychopathology, in consideration of possible contributions of other maternal psychopathology, than PTSD and depression to preschool children's behavior problems.

The primary purpose of this study was to assess the relative impact of maternal depression alone versus conjoined maternal depression and PTSD on child behavior problems. Future studies can extend these findings by comparing the effects of maternal PTSD alone with the relative effects of maternal depression alone. This study could not illuminate whether maternal depression existed prior to exposure to the WTC attacks. In addition, this study focused on maternal depression and PTSD as categorical variables and examined their association to clinically significant behavior problems in preschoolers (also a categorical variable). Thus, this study's focus addressed itself to the effects of clinically significant maternal psychological problems on clinically significant child behavioral problems. Future studies will address the associations of maternal PTSD and depression and preschool child behavioral problems using continuous measures of symptomatology and child behavior problems.

Studies of child and maternal vulnerability prior to exposure to terrorism and disaster would help clarify the interaction of vulnerability and exposure. Similarly, the design of this study could not address whether maternal PTSD alone might have been found more frequently closer in time to the actual exposure to the WTC attacks among these mothers of young children. Answering these types of questions would require the use of prospective longitudinal designs. In short, an important focus for future research should be to characterize the course of natural recovery of mother and child over time, ideally considering preexposure vulnerability, and examining dyadic interaction patterns. Finally, although these findings point to the importance of maternal psychopathology in the behavioral outcomes associated with terrorism exposure in preschool children, we cannot be certain that our results will generalize to other types of terrorism attacks (chronic and recurrent; other types of smaller impact single events) and to the effects of other disasters.

Clinical Implications

These data strongly suggest that very young children are vulnerable to the effects of terrorism and, given the time of assessment postexposure, that these behavioral effects are persistent. Yet, it appears that young children were provided services far less than other populations following the WTC attacks (Covell et al., 2006). We previously recommended that one strategy to increase efficacy in identifying children at need of services would be through screening by pediatricians (Chemtob et al., 2008).

The present findings suggest that a complementary strategy to identify very young children at risk is through identifying psychopathology among the mothers of very young children in the aftermath of large-scale disasters. Moreover, particular attention should be given to identifying mothers with co-occurring depression and PTSD when developing a strategy to identify very young children who are at risk subsequent to terrorism exposure. Given the apparent persistence of the effects of terrorism, interventions need to consider multiple stressors faced by both mothers and children over time (Appleyard & Osofsky, 2003; see also Chemtob et al., 2008, as recovery from exposure to terrorism occurs over time in the context of often challenging postdisaster recovery environments.

Screening and intervening with the mothers of young children may serve to help both mother and young child more efficaciously. In this regard, our findings add to the growing literature pointing to the importance of considering the mother–child dyad in conceptualizing the effects of trauma as well as its treatment especially when very young children are involved (Chemtob et al., 2008; Lieberman & Van Horn, 2005; Schechter et al., 2002). Promising data have indicated that dyadic intervention can enhance the well-being of both mother and young child following domestic violence exposure (Lieberman, Ghosh Ippen, & Van Horn, 2006). There is a need for dyadic interventions that target supporting young children and their mothers following exposure to terrorism and disaster. As well, Weissman has shown that treating maternal depression alone can have a salutary effect on child symptoms (Weissman et al., 2006). There is certainly value in pursuing both dyadic and mother-centered treatment research that focuses on ameliorating outcomes for mothers with conjoined PTSD and depression and their young children following exposure to terrorism and to other types of disaster.

References

Achenbach, T. M., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms & profiles.* Burlington: University of Vermont Press.

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles.* Burlington: University of Vermont Press.

Appleyard, K., & Osofsky, J. D. (2003). Parenting after trauma: Supporting parents and caregivers in the treatment of children impacted by violence. *Infant Mental Health Journal, 24,* 111–125.

Barkley, R. A., Shelton, T. L., Crosswait, C., Moorehouse, M., Fletcher, K., Barrett, S., et al. (2000). Multi-method psycho-educational intervention for preschool children with disruptive behavior: Preliminary results at post-treatment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 319–332.

Binder, D. A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika, 79,* 139–147.

Bollen, K. A., Glanville, J. L., & Stecklov, G. (2007). Socioeconomic status, permanent income, and fertility: A latent-variable approach. *Population Studies, 61,* 15–34.

Breslau, N. (1985). Depressive symptoms, major depression, and generalized anxiety—A comparison of self-reports on CES–D and results from diagnostic interviews. *Psychiatry Research, 15,* 219–229.

Breslau, N., Davis, G. C., & Prabucki, K. (1988). Depressed mothers as informants in family history research—Are they accurate? *Psychiatry Research, 24,* 345–359.

Calkins, S. D., & Dedmon, S. E. (2000). Physiologic and behavioral regulation in two-year-old children with aggressive/destructive behavior problems. *Journal of Abnormal Child Psychology, 28,* 103–118.

Chemtob, C. M., Nomura, Y., & Abramovitz, R. (2008). Impact of conjoined exposure to the World Trade Center attacks and to other traumatic events on the behavioral problems of preschool children. *Archives of Pediatrics and Adolescent Medicine, 162,* 126–132.

Chemtob, C. M., Roitblat, H. L., Hamada, R. S., Carlson, J. G., & Twentyman, C. T. (1988). A cognitive action theory of post-traumatic stress disorder. *Journal of Anxiety Disorders, 2,* 253–275.

Cohn, J. F., & Tronick, E. (1989). Specificity of infants response to mothers affective behavior. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 242–248.

Covell, N. H., Allen, G., Essock, S. M., Pease, E. A., Felton, C. J., Lanzara, C. B., et al. (2006). Service utilization and event reaction patterns among children who received Project Liberty counseling services. *Psychiatric Services, 57,* 1277–1282.

DeVoe, E. R., Bannon, W. M., & Klein, T. P. (2006). Post-9/11 helpseeking by New York City parents on behalf of highly exposed young children. *American Journal of Orthopsychiatry, 76,* 167–175.

Dosemeci, M., & Benichou, J. (1998). An alternative test for trend in exposure-response analysis. *Journal of*

Exposure Analysis and Environmental Epidemiology, 8, 9–15.

Downey, G., & Coyne, J. C. (1990). Children of depressed parents—An integrative review. *Psychological Bulletin, 108,* 50–76.

Fergusson, D. M., Lynskey, M. T., & Horwood, L. J. (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology, 21,* 245–269.

Foa, E. B. (1995). *Manual for the Posttraumatic Stress Diagnostic Scale (PDS).* Minneapolis, MN: NCS Pearson.

Fremont, W. P. (2004). Childhood reactions to terrorism-induced trauma: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 381–392.

Galea, S., Ahern, J., Resnick, H., Kilpatrick, D., Bucuvalas, M., Gold, J., et al. (2002). Psychological sequelae of the September 11 terrorist attacks in New York City. *New England Journal of Medicine, 346,* 982–987.

Ghosh Ippen, C., Ford, J., Racusin, R., Acker, M., Bosquet, M., Rogers, K., et al. (2002). *Traumatic Events Screening Inventory—Parent Report Revised.* San Francisco: The Child Trauma Research Project of the Early Trauma Network and the National Center for PTSD Dartmouth Child Trauma Research Group.

Green, B. L., Korol, M., Grace, M. C., Vary, M. G., Leonard, A. C., Gleser, G. C., et al. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 945–951.

Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J. M., Richardson, R., et al. (2006). The equivalence of the Child Behavior Checklist 1½-5 across parent race/ethnicity, income level, and language. *Psychological Assessment, 18,* 313–323.

Hakulinen, T. A. (1981). Mantel-Haenszel statistic for testing the association between a polychotomous exposure and a rare outcome. *American Journal of Epidemiology, 113,* 192–197.

Hirshfeld, D. R., Biederman, J., Brody, L., Faraone, S. V., & Rosebaum, J. F. (1997). Expressed emotion toward children with behavioral inhibition: Associations with maternal anxiety disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 367–399.

Kilic, E. Z., Ozguven, H. D., & Sayil, I. (2003). The psychological effects of parental mental health on children experiencing disaster: The experience of Bolu earthquake in Turkey. *Family Process, 42,* 485–495.

Kochanska, G., Kuczynski, L., Radke-Yarrow, M., & Welsh, J. D. (1987). Resolutions of control episodes between well and affectively ill mothers and their young-children. *Journal of Abnormal Child Psychology, 15,* 441–456.

Kotler, B. P., & Brown, J. S. (1990). Harvest rates of two species of gerbilline rodents. *Journal of Mammalogy, 71,* 591–596.

Laor, N., Wolmer, L., & Cohen, D. J. (2001). Mothers' functioning and children's symptoms 5 years after a SCUD missile attack. *American Journal of Psychiatry, 158,* 1020–1026.

Liberatos, P., Link, B. G., & Kelsey, J. L. (1988). The measurement of social class in epidemiology. *Epidemiologic Review, 10,* 87–121.

Lieberman, A. F., Ghosh Ippen, C., & Van Horn, P. (2006). Child-parent psychotherapy: 6-month follow-up of a randomized controlled trial. *Journal of American Academy of Child & Adolescent Psychiatry, 45,* 913–918.

Lieberman, A. F., & Van Horn, P. (2005). *Don't hit my mommy.* Washington, DC: Zero to Three Press.

Lin, D. Y., & Wey, L. J. (1989). The robust inference for the proportional hazards model. *Journal of American Statistical Association, 84,* 1074–1078.

McCloskey, L. A., & Walker, M. (2000). Posttraumatic stress in children exposed to family violence and single-event trauma. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 108–115.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine, 338,* 171–179.

Mesman, J., Bongers, I. L., & Koot, H. M. (2001). Preschool developmental pathways to preadolescent internalizing and externalizing problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 679–689.

Michael, T., & Ehlers, A. (2007). Enhanced perceptual priming for neutral stimuli occurring in a traumatic context: Two experimental investigations. *Behaviour Research and Therapy, 45,* 341–358.

Michael, T., Ehlers, A., & Halligan, S. L. (2005). Enhanced priming for trauma-related material in posttraumatic stress disorder. *Emotion, 5,* 103–112.

Murray, L. (1992). The impact of postnatal depression on infant development. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33,* 543–561.

Najarian, L. M., Goenjian, A. K., Pelcovitz, D., Mandel, F., & Najarian, B. (1996). Relocation after a disaster: Posttraumatic stress disorder in Armenia after the earthquake. *Journal of American Academy of Child and Adolescent Psychiatry, 35,* 347–383.

Olff, M., Langeland, W., Draijer, N., & Gersons, B. P. R. (2007). Gender differences in posttraumatic stress disorder. *Psychological Bulletin, 133,* 183–204.

Osofsky, J. D., & Eberhart-Wright, A. (1988). Affective exchanges between high risk mothers and infants. *International Journal of Psychoanalysis, 69,* 221–231.

Pfefferbaum, B. J. (2005). Psychological impact of terrorism on children and families in the United States. *Journal of Aggression, Maltreatment and Trauma, 9,* 305–317.

Pine, D. S., Costello, J., & Masten, A. (2005). Trauma, proximity, and developmental psychopathology: The effects of war and terrorism on children. *Neuropsychopharmacology, 30,* 1781–1792.

Radloff, L. S. (1977). The CES–D Scale: A self-report depression scale for research in the general population. *Journal of Applied Psychological Measurement, 1,* 385–401.

Renouf, A. G., & Kovacs, M. (1994). Concordance between mothers' reports and children's self-reports of depressive symptoms—A longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*, 208–216.

Richters, J., & Pellegrini, D. (1989). Depressed mothers' judgments about their children: An examination of the depression-distortion hypothesis. *Child Development, 60*, 1068–1075.

Schechter, D. S., Zeanah, C. H., Myers, M. M., Brunelli, S., Coates, S. W., Grienenberger, J., et al. (2002, July). *Negative and distorted maternal attributions among violence-exposed mothers of very young children before and after single-session videofeedback: Are maternal psychopathology and reflective functioning predictive?* Paper presented at the meetings of the World Association of Infant Mental Health, Amsterdam, Netherlands.

Schmidt, K. A. (2000). Interactions between food chemistry and predation risk in fox squirrels—Statistical data included. *Ecology, 29*, 1–13.

Shaw, D. S., Owens, E. B., Giovannelli, J., & Winslow, E. B. (2001). Infant and toddler pathways leading to early externalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 36–43.

Smith, P., Perrin, S., Yule, W., & Rabe-Hesketh, S. (2001). War exposure and maternal reactions in the psychological adjustment of children from Bosnia-Hercegovina. *Journal of Child Psychology and Psychiatry, 42*, 395–404.

Stein, A., Gath, D. H., Bucher, J., Bond, A., Day, A., & Cooper, P. J. (1991). The relationship between postnatal depression and mother-child interaction. *British Journal of Psychiatry, 158*, 46–52.

Thabet, A. A. M., Karim, K., & Vostanis, P. (2006). Trauma exposure in pre-school children in a war zone. *British Journal of Psychiatry, 188*, 154–158.

Vila, G., Witkowski, P., Tondini, M. C., Perez-Diaz, F., Mouren-Simeoni, M. C., & Jouvent, R. (2001). A study of posttraumatic disorders in children who experienced an industrial disaster in the Briey region. *European Child and Adolescent Psychiatry, 10*, 10–18.

Vythilingam, M., Blair, K. S., McCaffrey, D., Scaramozza, M., Jones, M., Nakic, M., et al. (2007). Biased emotional attention in post-traumatic stress disorder: A help as well as a hindrance? *Psychological Medicine, 37*, 1445–1455.

Weissman, M. M., Pilowsky, D. J., Wickramaratne, P. J., Talati, A., Wisniewski, S. R., Fava, M., et al. (2006) Remissions in maternal depression and child psychopathology—A STAR*D-child report. *Journal of the American Medical Association, 295*, 1389–1398.

Wolmer, L., Laor, N., Gershon, A., Mayes, L. C., & Cohen, D. J. (2000). The mother-child dyad facing trauma—A developmental outlook. *Journal of Nervous and Mental Disease, 188*, 409–415.

Yehuda, R. (2002). Clinical relevance of biologic findings in PTSD. *Psychiatric Quarterly, 73*, 123–133.

Yehuda, R., & Bierer, L. M. (2008). Transgenerational transmission of cortisol and PTSD risk. *Progress in Brain Research, 167*, 121–135.

Zahn-Waxler, C., Iannotti, S., Cummings, E., & Denham, S. (1990). Antecedents of problem behaviors in children of depressed mothers. *Development and Psychopathology, 2*, 271–291.

Zhang, J., & Yu, K. F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association, 280*, 1690–1691.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1142–1160

Exposure to 9/11 Among Youth and Their Mothers in New York City: Enduring Associations With Mental Health and Sociopolitical Attitudes

Elizabeth T. Gershoff
University of Texas at Austin

J. Lawrence Aber
New York University

Angelica Ware
Austin Independent School District

Jennifer A. Kotler
Sesame Workshop

The enduring impact of exposure to the 9/11 terrorist attacks on mental health and sociopolitical attitudes was examined in a sample of 427 adolescents (M = 16.20 years) and their mothers residing in New York City. Direct exposure to the terrorist attack was associated with youth depression symptoms and with mothers' posttraumatic stress disorder symptoms. There was no evidence of reciprocal effects of mother exposure on youth or of youth exposure on mothers. Although mothers reported engaging in more emotional processing coping assistance with their children, coping assistance was not associated with youth's symptomatology. Media exposure was found to be a strong predictor of youth's and mothers' sociopolitical attitudes about issues such as prejudice toward immigrants, social mistrust, and current events.

It is not an exaggeration to say that nearly everyone in the United States was affected in some way by the September 11, 2001, terrorist attacks on New York City (NYC) and Washington, DC. The dramatic events of 9/11 were reasonably expected to have far-reaching and long-lasting effects on the mental health of adults and children throughout the United States. Yet, undoubtedly those individuals who witnessed the attacks and their aftermath first-hand will have been especially affected by the event. Without any form of warning, lower Manhattan was transformed into a disaster zone when terrorists attacked the World Trade Center (WTC) on 9/11. The scale of the devastation and the large number of victims meant that residents throughout NYC were directly affected. Individuals throughout the city had direct exposure to the 9/11 attacks by physically being at or near the WTC, by seeing the towers in flames or collapsing, by smelling the cloud of smoke that hovered over the city for days, or by having their day severely disrupted by lack of subway and bus transportation or by the pandemonium that waged throughout the city as residents worried if another attack was imminent and tried to reach their loved ones through clogged phone lines. Indeed, in the months following the attack, New Yorkers initially exhibited high rates of posttraumatic stress disorder (PTSD), with estimated rates of between 7.5% and 11.2% of probable PTSD among adult residents (Galea et al., 2002; Schlenger et al., 2002) compared to a rate of 5.8% among a national sample of Americans 6 months after the attacks (Silver, Holman, McIntosh, Poulin, & Gil-Rivas, 2002). A study of a representative sample of NYC school children 6 months after 9/11 determined that 28% of the students had one or more anxiety or depressive disorders and that

---

We wish to thank the collaboration among the New York City Department of Education, Applied Research Consulting LLC, Columbia University Mailman School of Public Health, and New York State Psychiatric Institute for allowing us to use their measure of the impact of 9/11 on New York City public school children. We also wish to thank Irene Strum and Henry Solomon at the New York City Department of Education for their assistance in maintaining contact with this longitudinal sample, and to thank Jim Mercy and Rodney Hammond in the Division of Violence Prevention at the Centers for Disease Control and Prevention for helping secure supplemental funding for a module of questions related to 9/11. This work was conducted while the authors were employed at the National Center for Children in Poverty, Mailman School of Public Health, Columbia University, and was funded through grants from the Centers for Disease Control and Prevention through the Columbia Center for Youth Violence Prevention (CCR218598) and from the National Institute of Mental Health (1R01MH63685) awarded to the first and second authors.

Correspondence concerning this article should be addressed to Elizabeth T. Gershoff, Department of Human Development and Family Sciences, University of Texas at Austin, 1 University Station A2702, Austin, TX 78712. Electronic mail may be sent to liz.gershoff@austin.utexas.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0009

greater exposure to 9/11 was associated with greater levels of problematic mental health symptoms (Hoven et al., 2005). In this article, we examine whether exposure to the 9/11 attacks had enduring effects on a sample of youth and their mothers residing in NYC.

Parents are typically more negatively affected by a disaster than are nonparent adults (Norris, Friedman, Watson, Byrne, et al., 2002) and thus we would expect New Yorkers who were parents at the time of the 9/11 attacks to have been at especially heightened risk for mental health problems. Parents are at increased risk for problems because the difficulties they face in addressing their own anxieties after a disaster are compounded by their empathy for their children's distress as well as their responsibility to help their children cope with the disaster. Phillips, Featherman, and Liu (2004) documented what they termed the evocative influence of children's stress symptoms on parents' own feelings of shaken security and threat, with adults exposed to children distressed by the events of 9/11 being twice as likely to feel threatened than adults who did not encounter distressed children. However, very little is known about how parents' mental health was affected by their children's exposure to the 9/11 attack, over and above their own exposure. We explore the possibility of reciprocal effects of parent and youth exposure on each others' mental health problems in this article.

Youth often turn to their parents for support in dealing with a traumatic event. In the aftermath of 9/11, 85% of youth residing just outside NYC used their parents as a source of support in the week after the 9/11 attacks, and 71% found parents' support to be beneficial (Gould, Munfakh, Kleinman, Lubell, & Provenzano, 2004). Although children's strategies for coping with a disaster or other stressful event have been shown to predict improved functioning (Compas, Worsham, & Ey, 1992; Vernberg, La Greca, Silverman, & Prinstein, 1996), and children do report receiving the most support from their parents and friends in the wake of disasters such as Hurricane Andrew or 9/11 (Phillips, Prince, & Schiebelhut, 2004; Prinstein, La Greca, Vernberg, & Silverman, 1996), there has been only limited research on the role parents play in helping children select coping strategies. This is despite the fact that official advice to parents in the wake of disasters specifically encourages parents to assist their children's coping with the disaster by talking with children and listening to their concerns as well as returning to predisaster routines as soon as possible (Federal Emergency Management Agency, 2004). Those studies of which we are aware have used child reports of parents' coping assistance (Prinstein et al., 1996; Vigil & Geary, 2008). In contrast, we examine parents' self-reports of their coping assistance, first as predicted by their own or their youth's exposure to a disaster, and then as predictors of youth mental health problems.

The effects of exposure to man-made disasters such as 9/11 are not restricted to mental health. Sociopolitical attitudes, such as prejudice toward immigrants and social mistrust, could also be affected. In particular, civil liberties groups were concerned that the nature of the attacks on NYC and Washington might evoke feelings of resentment of or prejudice toward immigrants and general mistrust of others. Such fears appear to have been unfounded; rather, the opposite appears to have occurred, with youth reporting more interest in politics and civic issues post-9/11 (Metz & Youniss, 2003). Putnam (2002) reported that young Americans' interest in public affairs increased by 27%, whereas trust in community leaders grew by 19%. Neither was there an observed increase in negative attitudes about immigrants as a result of the 9/11 attacks; even in the weeks immediately after 9/11, the majority (87%) of respondents in a national survey agreed that immigrants help make the United States more open to new cultures and ideas, and only 24% believed that immigrants contribute to increased crime in the United States (Traugott et al., 2002). In a survey 2 weeks after the attack, 61% of New Yorkers agreed that people are fair, 71% agreed that people are helpful, and 34% agreed that people are trustworthy; these attitudes remained largely unchanged 5 months after the attacks (Rasinski, Berktold, Smith, & Albertson, 2002). In this study, we examined whether and how the amount and type of individual and partner exposure to the events of 9/11 were associated with NYC youth's and mothers' sociopolitical attitudes several months after the attack.

The unpredictable nature of disasters such as terrorist attacks makes it especially difficult to judge their impact on individuals. Studies already in the field, preferably ones close to the disaster site, can provide a quick venue for disaster assessment by "piggybacking" disaster assessments onto preexisting studies (Gershoff & Aber, 2004). The results reported here draw from a study employing such a strategy. The study is the result of an amendment to a planned long-term follow-up of children in an evaluation of one of the largest school-based violence prevention programs in the country, the Resolving Conflict Creatively Program (RCCP).

The influence of RCCP on the social-cognitive, behavioral, and academic development of elementary school children has been reported elsewhere by Aber and colleagues (Aber, Brown, & Jones, 2003; Aber, Jones, Brown, Chaudry, & Samples, 1998; Brown, Roderick, Lantieri, & Aber, 2004). We include measures of exposure to the RCCP intervention as controls in these analyses.

Data collection began 4 months after 9/11 and ended 2 years later; interviews were on average 15 months after 9/11. This long time lapse between 9/11 and data collection is not ideal for detecting immediate effects. However, findings that individuals exposed to disasters tend to show impaired functioning for a full year after the event, and that for some individuals these effects last for years (Norris, Friedman, & Watson, 2002), suggest that our time window is appropriate for assessing lasting impacts of 9/11 exposure among individuals residing in NYC.

The first aim of this study was descriptive, namely to document the rates of exposure to the 9/11 terrorist attacks on the WTC among a sample of youth and their mothers in NYC. Unlike studies that have used a pre- and postattack design that assumes equal exposure to the attacks and their aftermath (e.g., Hendricks & Bornstein, 2007; Henry, Tolan, & Gorman-Smith, 2004; Metz & Youniss, 2003), we assessed three forms of actual exposure to the events of 9/11: individual direct exposure to the attacks and their immediate aftermath, family or friend direct exposure at the WTC site, and media exposure to coverage of the attack. To understand whether some youth and mothers were more likely than others to be exposed to the attack, we examined several demographic characteristics that have been associated with how well individuals cope with a disaster, namely gender, age, ethnicity, and socioeconomic status (see Norris, Friedman, Watson, Byrne, et al., 2002, and La Greca, Silverman, Vernberg, & Roberts, 2002, for reviews). We also were interested in whether youth and mothers reported similar or distinct rates of the three forms of exposure, given that youth tend to be more adversely affected by disasters than adults (see Norris, Friedman, Watson, Byrne, et al., 2002).

To determine the potential impact of 9/11 exposure on mental health, our second aim was to establish whether, controlling for demographic characteristics associated with exposure, self-reported direct, family and friend, and media exposure were associated with increases in youth's and mothers' PTSD and depression symptoms. We chose to focus on PTSD and depression because they are the two mental health problems most often observed after disasters (La Greca et al., 2002; Norris, Friedman, Watson, Byrne, et al., 2002). To examine potential bidirectional effects, we also tested whether partner exposure (i.e., mother exposure predicting youth symptoms, and youth exposure predicting mother symptoms) added to the likelihood that youth or their mothers reported symptoms of PTSD or depression.

Our third aim was to understand the role of parents' coping assistance in the face of such a disaster. We sought to determine whether the extent to which mothers assisted their adolescents in coping with the events of 9/11 was predicted by their own exposure or by their adolescents' exposure. We then examined whether mothers' coping assistance predicted lower rates of youth PTSD and depression symptomatology.

Finally, our fourth aim was to establish whether exposure to the events of 9/11 had any enduring association with youth's and mothers' reports of their families' and their personal sociopolitical attitudes. Specifically, we considered families' socialization around prejudice, social mistrust, and current events, as well as individuals' own attitudes about the level of trust in their communities and about immigrants' roles in their communities. As with the prediction of mental health, we also examined whether partner exposure predicted participants' sociopolitical attitudes over and above their own exposure.

Method

Participants

The sample for this study was drawn from a larger study of 908 youth who participated in a longitudinal follow-up of children in an evaluation of one of the largest school-based violence prevention programs in the country, the RCCP. For 498 of these youth, their primary caregiver was also interviewed. Of these 498 families, 31 were excluded because they were missing adolescent or caregiver exposure to the events of 9/11. Finally, only 40 of the primary caregivers were male; they were excluded from the analyses.

The final analysis sample consisted of 427 adolescents and their female caregivers residing in NYC. Because the vast majority of the female caregivers were mothers or stepmothers to the youth (94%), we will subsequently refer to them as "mothers." Adolescents ranged in age from 12 to 20 years ($M = 16.20$, $SD = 1.69$). The youth sample was 54% females and 47% Hispanic American, 35% African American, 14% European American, and 4% of another race (e.g., Asian, Native American). The majority of the adolescents reported the United States as their country of origin (87%), but nearly two thirds (64%) of them had at least one parent born outside the United States. In the follow-up, 31% of the participants were in 7th–9th grades, 54% were in 10th–12th grades, 5% were in a GED (general equivalency diploma) or vocational program, 6% were in college, 2% were working and not in school, and 2% were not working or not in school. Of the students in school, 26% reported that they worked part-time or full-time as well. Most adolescents indicated that they attended or had attended a public junior or senior high school (91%) in NYC.

The mothers averaged 42 years of age (*SD* = 7.11). Half of the mothers reported that they were married (50%). The race-ethnic breakdown of the mothers was roughly equal to that of the adolescents: 44% Hispanic American, 37% African American, 15% European American, and 4% of another race. Two thirds of the mothers worked (63%). The mothers reported a range of educational backgrounds: Ten percent had only elementary school education; 17% had attended, but not graduated from, high school; 28% were high school graduates; 26% had attended college for 1–3 years; and 19% were college graduates. The median education level among mothers was high school graduate. Fifty-one percent of the mothers were born outside the United States, in a total of 35 countries. Of the foreign countries of origin, the most frequently reported were the Dominican Republic (*n* = 50), Puerto Rico (*n* = 37), Jamaica (*n* = 23), and Haiti (*n* = 20).

Protocol

Families were recruited for the study through mailings and phone calls using contact information obtained from the New York City Department of Education. Recruitment letters were written in both English and Spanish. Once consent was obtained from parents and assent or consent was obtained from youth (depending on whether they were under or over 18), youth were interviewed in person by trained research assistants for 2–3 hr using a structured format. Youth entered their responses directly into a laptop computer. To minimize any intimidation adolescents might feel, the interviewers were diverse (over half of the interviewers were racial-ethnic minorities), primarily female, and entirely young adults. The interviews primarily took place at public spaces such as community libraries or local parks, and commercial establishments such as coffee shops, or, less frequently, adolescents' homes. The students were compensated $50 for their time.

Mother interviews followed a structured format that lasted approximately 1 hr. The majority of mother interviews were conducted by reading the complete questionnaire over the phone (55%). In 34% of the cases, the parent filled out a paper copy of the questionnaire in person with the interviewer, whereas 11% of the mothers filled out the paper copy on their own and mailed it to us. Both the written and oral versions of the parent interview were available in Spanish for mothers who preferred to read and speak Spanish; 21% of the mothers were either read the Spanish version of the questionnaire or completed a paper copy of the Spanish version. Mothers were compensated $25 for their time.

Measures

Exposure to the 9/11 Attacks on the WTC

Exposure to the terrorist attacks on NYC on September 11, 2001, and their aftermath were assessed through sections of an instrument created by Hoven et al. (2002, 2005) through collaboration with the New York City Department of Education and the Centers for Disease Control and Prevention. Three forms of exposure to the events of 9/11 were assessed: direct exposure, media exposure, and family and friend exposure. *Direct exposure* was a count of 13 potential experiences with the event (e.g., seeing the planes crash or the towers collapse, having to leave where they were for safety concerns, smelling the smoke after the event; see Table 1 for all items). To assess *family and friend exposure*, we asked respondents whether they knew someone who escaped the WTC unharmed, someone who was at the WTC on 9/11 and was physically hurt, or someone who was killed in the WTC. The family and friend exposure scale was a count across these three categories and thus ranged from 0 to 3. We asked youth and mothers to report the extent of their *media exposure* to the events of 9/11 by asking them how much they reported learning about the event from (a) TV; (b) newspapers, radio, or magazines; or (c) the Internet, on a scale of 0 (*none*), 1 (*some*), or 2 (*a lot*). Each respondent's media exposure score was an average of these three ratings.

Youth interviews were conducted an average of 443 days (*SD* = 176) after September 11, 2001, which is roughly equivalent to 15 months. Mothers

Table 1
Youth's and Mothers' Direct Exposure to the 9/11 Attack on the World Trade Center (WTC)

Type of direct exposure	Youth (%)	Mothers (%)
Smelled the smoke from the WTC buildings after the attack	51	64
Someone in family works near the WTC site	35	35
Had difficulty getting home that day	34	33
Had to leave the location where they were for safety reasons	29	18
Was in or near the cloud of smoke or dust caused by the attack	20	25
Parents have cut down on freedom to travel around the city since September 11 (youth)/ Feel travel around city is limited because of the attack (mothers)	18	53
Someone in family has lost job because of the attack	15	20
Saw the planes crash or the towers collapse with own eyes	10	12
Someone in family worked in rescue or relief services at the WTC site	12	19
Had to attend another school/workplace temporarily because it was closed due to the attack	1	6
Changed school/job permanently because of the attack	1	4
Was physically hurt during the attack	1	3
Had to move out of home because of the attack	< 1	2
Any one of the above	87	91

were interviewed on average 477 days (*SD* = 167) after 9/11, which is also roughly equivalent to 15 months. We included the intervals between 9/11 and the mothers' and youth's interviews as controls in all analyses.

Mothers' Coping Assistance to Youth

Mothers were asked to report the kinds of assistance they provided to their children in reaction to the 9/11 attacks using a version of the Children's Coping Assistance Checklist that we adapted to be a parent-report measure (Prinstein et al., 1996). Three types of coping assistance were assessed: emotional processing, roles and routines, and distraction. *Emotional processing* included items such as, "I talked with my child about things she or he remembered most about the attack." The *roles and routines* subscale included items such as, "I did things with my child that made him or her feel like things were back to normal since the attack." The *distraction* coping assistance subscale included items such as, "I did things with my child to help him or her forget about the attack." In this study, the coping assistance subscales had strong internal consistency: emotional processing, $\alpha = .90$; roles and routines, $\alpha = .84$; distraction, $\alpha = .90$. Each subscale ranged from 1 to 4 (emotional processing: $M = 2.36$, $SD = 0.71$; roles and routines: $M = 2.66$, $SD = 0.84$; distraction: $M = 2.25$, $SD = 0.95$).

Youth's Mental Health

Youth's *PTSD symptoms* and *depression symptoms* were both assessed using the Computer Diagnostic Interview Schedule for Children–IV (C-DISC Development Group, 2000). The C-DISC is a highly structured interview designed to assess *DSM–IV* psychiatric disorders and symptoms in children and adolescents aged 9–17 years through self-report. (Although a young adult version for 18- to 25-year-olds was under development and not available for use in this study, the DISC has been used successfully with an 18- to 24-year-old population; see Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000.) The adolescents were asked if they had experienced symptoms of major depression or PTSD during the past year. Previous research has found that the C-DISC–IV is answered more consistently than any other psychiatric diagnostic interview that has been prepared for either children or adults and has established reliability (C-DISC Development Group, 2000; Shaffer et al., 2000). Although clinical cutoffs applied by the DISC algorithm revealed that 4% of youth reached the level of a clinical diagnosis of PTSD symptoms in the last year and that 12% reached clinical levels of major depression disorder, for these analyses we used continuous counts of symptoms to capture the full range of symptomatology. It is also important to note that the PTSD symptoms reported by adolescents could be about any traumatic life event, not necessarily the events of September 11; only 10 of the 427 youth reported September 11 as their major traumatic life event.

Mothers' Mental Health

Mothers' *PTSD symptoms* were assessed using the PTSD subscale of the DISC Predictive Scales (DPS; Lucas et al., 2001). The DPS was created by the developers of the full interview to maximize specificity and sensitivity and to be predictive of full DISC diagnoses (Lucas et al., 2001). The DPS PTSD subscale is an eight-item scale; scores range from 0 to 1. For this sample of mothers, the DPS PTSD scale had acceptable internal consistency, $\alpha = .74$ ($M = 0.34$, $SD = 0.25$). The developers of the DPS suggest that the endorsement of five or more symptoms indicates a probably clinical level of the disorder (Lucas et al., 2001); applying this criterion, 20% of the mothers in our sample had probable clinical levels of PTSD. However, for our analyses, we used the symptom count as a continuous score to best capture the full variability of PTSD symptoms among mothers in our sample.

Mothers' *depression symptoms* were identified using the Center for Epidemiological Studies Depression Scale (CES–D; Radloff, 1977). Designed to assess depressive symptoms in normal populations, this widely used scale asks respondents to rate the frequency with which 20 states or behaviors related to depression occurred during the past week (e.g., "I felt sad," "I felt that everything I did was an effort," "I enjoyed life"). The CES–D scale had excellent internal consistency with this sample ($\alpha = .90$; $M = 12.65$, $SD = 10.87$). Radloff (1977) identified 16 or more symptoms as a cutoff for potential clinical levels of depression. Mothers in our sample had CES–D scores of as high as 57 (of a possible 60), and 34% of the mothers reported elevated depression symptoms at or above the cutoff score of 16. As with youth's mental health symptoms and mothers' PTSD symptoms, however, we used the count of depressive symptoms as a continuous measure in our analyses.

Family and Personal Sociopolitical Attitudes

Both youth's and mothers' perceptions of the mothers' socialization around certain sociopolitical issues and of their own sociopolitical attitudes were assessed using the Intergroup Understanding or Bias Questionnaire (Flanagan, Gill, & Gallay, 1998; Flanagan, Ingram, Gallay, & Gallay, 1997). The questionnaire was written for adolescents and young adults; we created a companion version for parents. For this study, we used three subscales from this measure regarding sociopolitical socialization: family lessons about prejudice, family discussions of current events, and family lessons about social mistrust. We also asked youth and mothers to complete subscales tapping two aspects of their sociopolitical attitudes, namely social integration and prejudice toward immigrants. Responses to all items were along a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

Family lessons about prejudice. This subscale included five items that asked respondents how much they agreed with statements such as, "My parents have taught me [I have taught my children] it is wrong to judge people before you get to know them" and "My parents have told me [I have told my children] there might be times when people will be prejudiced against me [them]." This subscale showed strong internal consistency for the youth ($\alpha = .78$; $M = 4.07$, $SD = 0.74$) and adequate internal consistency for the mothers ($\alpha = .69$; $M = 4.38$, $SD = 0.54$) in this sample.

Family lessons about social mistrust. This subscale consisted of four items, including, "My parents have warned me [I have warned my children] that people sometimes take advantage of you," and "My parents have warned me [I have warned my children] you can't always trust people." This subscale had strong internal consistency for the youth ($\alpha = .83$; $M = 4.21$, $SD = 0.77$) and adequate consistency for the mothers ($\alpha = .65$; $M = 4.33$, $SD = 0.52$) in this sample.

Family discussions of current events. Comprising six items, this subscale includes items such as, "In my house we have many discussions about current events and politics" and "I tell my parents [my children] my opinions about events in the news." This subscale showed strong internal consistency for the youth ($\alpha = .83$; $M = 3.36$, $SD = 0.81$) and adequate internal consistency for the mothers ($\alpha = .71$; $M = 3.77$, $SD = 0.60$) in this sample.

Social integration. The six items in this subscale include statements such as, "Most people in our community try to make this a good place to live" and "If someone in our community has a problem, they can usually count on others to help them out." This subscale showed equally strong internal consistency for the youth ($\alpha = .84$; $M = 3.00$, $SD = 0.81$) and the mothers ($\alpha = .84$; $M = 3.13$, $SD = 0.76$).

Prejudice toward immigrants. For this subscale, respondents indicated their agreement with nine statements such as, "If they want to be accepted, immigrants should learn our customs and language" and "The children of immigrants should have the right to a good education in our country" (reversed). This subscale also showed strong internal consistency for both the youth ($\alpha = .78$; $M = 2.25$, $SD = 0.61$) and the mothers ($\alpha = .75$; $M = 3.65$, $SD = 0.56$) in this sample.

Components of RCCP Intervention

Experience with the intervention in middle childhood was operationalized using data on two primary RCCP components. *Classroom instruction* in RCCP is composed of the total number of lessons given by trained teachers to children in their classrooms across the 2 years of the evaluation (range = 0–10, *M* = 2.36, *SD* = 2.53). *Teacher training and coaching* is a count of the number of contacts a teacher had with the RCCP and consists of training sessions attended, one-on-one meetings with an RCCP staff developer, and classroom visits by the staff developer (range = 0–8, *M* = 2.10, *SD* = 2.02). (Please see Aber et al., 1998, 2003, for more in-depth discussions of these variables.) As stated earlier, we included these two aspects of the RCCP intervention in the analyses as controls only.

Results

Youth and Mother Exposure to the 9/11 Terrorist Attacks on the WTC

The vast majority of both youth (87%) and their mothers (91%) reported one or more forms of direct exposure to the 9/11 attacks on the WTC and their aftermath (see Table 1). Youth reported an average of 3.27 forms of direct exposure (*SD* = 1.85), with a low of 0 and a high of 9 forms of direct exposure. Mothers reported more forms of direct exposure, averaging 4.22 (*SD* = 2.29) and ranging from 0 to 12 forms of direct exposure. One in 10 youth and 1 in 8 mothers reported having seen the planes crash into the Twin Towers or the towers collapse with their own eyes. Thankfully, few reported having been physically hurt themselves in the attacks (youth: 1%; mothers: 3%), but 1 in 5 youth and 1 in 4 mothers reported being in or near the cloud of smoke or dust that resulted from the towers' collapse. One half of youth and nearly two thirds of mothers reported smelling the smoke from the WTC site after the attack. Equivalent numbers of youth and mothers reported having difficulty getting home that day (34% and 33%), although more youth reported that they had to leave the location where they were for safety reasons because of the attacks (youth: 29%; mothers: 18%). Half of the mothers felt that their travel around the city was limited because of the attacks (53%); we did not ask this specific question of the youth, but when we asked youth if they felt their parents had restricted their travel around the city because of the attack, only 18% said "yes." Youth and mothers were equally likely to say that someone in their family works near the WTC site (35%), although mothers were slightly more likely to report that someone in the family lost their job because of the attacks (20% vs. 15%) or that someone in the family worked in rescue or relief services at the WTC site (19% vs. 12%). Very few respondents (6% or less) reported that they had to change their school or their workplace, either temporarily or permanently, or that they had to move out of their home because of the attack.

Mothers were more likely than youth to report each form of family and friend exposure to the 9/11 attacks (see Table 2). Slightly less than half of all youth and two thirds of mothers reported one or more forms of family and friend exposure to 9/11; the average for youth was less than one form of family and friend exposure (*M* = 0.61, *SD* = 0.74), whereas for mothers the average was one form of family and friend exposure (*M* = 1.02, *SD* = 0.94). Mothers were twice as likely as youth to disclose that they knew someone who was killed in the attacks on the WTC (39% vs. 21%).

Regarding media exposure to 9/11, nearly all youth and all mothers reported learning about the attacks by watching TV (96% of youth and 98% of mothers; see Table 3). The percentages of youth and mothers who reported learning about the attacks from radio, newspapers, or magazines were also roughly equal at 93% and 91%, respectively. Youth were more likely to have used the Internet to learn about the attacks (38%) than mothers (25%). When their ratings for each of these forms of media exposure were combined into one subscale, youth's and mothers' average media exposures were strikingly similar: youth media exposure, *M* = 1.20, *SD* = 0.44; mothers' media exposure, *M* = 1.29, *SD* = 0.40.

Table 2
Youth's and Mothers' Family and Friend Exposure to the 9/11 Attack on the World Trade Center (WTC)

	Youth (%)	Mothers (%)
Knew someone who was in the WTC on the day of the attack but escaped without being hurt	34	47
Knew someone who was physically hurt in the WTC attack but not killed	7	16
Knew someone who was killed in the WTC	21	39
Any one of the above	47	64

Table 3
Youth's and Mothers' Media Exposure to the 9/11 Attack on the World Trade Center

Time spent learning about the attack from . . .	Youth (%)	Mothers (%)
TV		
A lot	69	87
Some	28	11
None	4	2
Radio, newspapers or magazines		
A lot	52	65
Some	41	26
None	7	9
Websites		
A lot	12	8
Some	27	17
None	62	75

We next examined whether certain groups of individuals were more vulnerable to exposure to the terrorist attacks than others. Table 4 contains the results of six separate regressions predicting youth and then mother direct, family and friend, and media exposure. Demographic characteristics largely did not predict exposure, with only two of the six models significant. Even the significant models, predicting youth's and then mothers' family and friend exposure, did only a modest job predicting the variance in exposure (8% and 10%, respectively). In the model predicting youth family and friend exposure, girls reported more, and African Americans and Hispanics Americans less, exposure. African Americans and Hispanic American mothers as well as mothers of another race or ethnicity also reported less family and friend exposure than European American mothers. Mothers with higher levels of education reported more family and friend exposure. Notably, the time lapse between 9/11 and the date of the youth or mother interview was not significantly associated with the amount of exposure reported by participants.

Exposure to 9/11 and Mental Health

To determine the impact of exposure to 9/11 on youth's and mothers' PTSD and depression symptoms, we conducted two regressions for each outcome. We first regressed the mental health outcome on youth and mother demographic characteristics and respondents' own 9/11 exposures to the model (e.g., youth exposure predicting youth PTSD symptoms), reported as Model A in Table 5. We then ran

Table 4
Demographic Predictors of Youth and Mother Exposure to 9/11

	Youth exposure to 9/11			Mother exposure to 9/11		
	Direct	Family and friend	Media	Direct	Family and friend	Media
Adolescent's age	.02	−.03	−.02	.09	.01	.10
Adolescent's gender[a]	.17***	.15**	.08	.04	.02	.00
Adolescent is immigrant	−.04	−.08	−.01	.05	−.07	.08
One or both parents are immigrants	.04	.05	.06	−.04	−.03	−.01
Is African American[b]	−.24**	−.31***	.03	−.09	−.24**	.17*
Is Hispanic American[b]	−.20**	−.35***	.10	−.13	−.23*	.14
Is of another race or ethnicity[b]	−.09	−.08	.06	−.05	−.15*	.04
Mother's age	−.03	.01	.02	−.03	−.03	.08
Mother is married	−.05	.02	.06	−.07	.06	.08
Mother's level of education	−.02	−.02	−.04	−.01	.11*	.06
Mother is employed	.04	−.06	.05	.09	.06	.04
Time between 9/11 and the interview[b]	−.01	−.00	.09	−.07	.08	−.09
RCCP intervention—Lessons	.08	.02	.02	.06	.10	.08
RCCP intervention—Staff development	−.09	−.01	−.03	−.09	−.09	−.15*
R^2	.06	.08	.04	.04	.10	.05
F	1.68	2.31**	1.14	1.15	2.96***	1.36
df	14, 373	14, 373	14, 373	14, 369	14, 369	14, 369

[a] 0 = *male*, 1 = *female*. [b] These characteristics were for youth in the models predicting youth exposure, and for parents in the models predicting parent exposure. European Americans are the reference category for the race-ethnicity variables.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5

Youth's and Mothers' PTSD and Depression Symptoms Predicted From Their Own and Each Others' Exposure to 9/11

	Youth's PTSD symptoms		Youth's depression symptoms		Mothers' PTSD symptoms		Mothers' depression symptoms	
	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)
Adolescent's age	.08	.07	.04	.04	.01	.01	.03	.02
Adolescent's gender^a	.07	.07	.15**	.15**	−.01	−.01	−.09	−.08
Adolescent is immigrant	−.09	−.10	−.17**	−.17**	.05	.04	.00	−.00
One or both parents are immigrants	−.13*	−.13*	−.04	−.04	.06	.06	−.09	−.09
Is African American^b	.24**	.22*	−.03	−.04	−.09	−.08	.06	.05
Is Hispanic American^b	.28**	.27**	−.04	−.04	.04	.06	.03	.03
Is of another race or ethnicity^b	.16**	.16*	−.03	−.03	−.02	−.01	.01	.01
Mother's age	.08	.08	.03	.03	−.07	−.06	−.09	−.09
Mother is married	−.02	−.02	.03	.04	−.10	−.09	−.11	−.10
Mother's level of education	−.04	−.04	−.08	−.08	−.23***	−.23***	−.22***	−.23***
Mother is employed	.07	.06	.02	.02	−.08	−.07	−.13*	−.12*
Time between 9/11 and the interview^b	−.08	−.07	−.02	−.01	.01	.02	−.01	.00
RCCP intervention—Lessons	.04	.04	.05	.05	.08	.08	.06	.06
RCCP intervention—Staff development	.00	.01	.02	.02	−.12	−.12	−.04	−.04
Youth's exposure to 9/11								
Direct exposure	.09	.08	.12*	.12*	—	.03	—	−.04
Family exposure^c	.01	.01	−.02	−.02	—	.02	—	.02
Media exposure	.06	.05	.03	.03	—	−.13*	—	−.06
Mothers' exposure to 9/11								
Direct exposure	—	.08	—	.02	.27***	.27***	−.02	−.01
Family exposure	—	−.02	—	−.04	.01	−.00	−.04	−.05
Media exposure	—	.05	—	−.01	−.01	.01	−.01	.00
R^2	.08	.09	.08	.08	.21	.23	.12	.13
F	1.83*	1.68*	1.86*	1.59*	5.04***	4.63***	2.62***	2.33***
df	17, 365	20, 362	17, 365	20, 362	17, 321	20, 318	17, 321	20, 318
R^2_Δ	—	.01	—	.00	—	.02	—	.01
F_Δ	—	0.85	—	0.16	—	2.00	—	0.71
df_Δ	—	3	—	3	—	3	—	3

Note. PTSD = posttraumatic stress disorder; RCCP = Resolving Conflict Creatively Program.

[a] 0 = *male*, 1 = *female*. [b] These characteristics were for youth in the models predicting youth mental health, and for parents in the models predicting parent mental health. European Americans are the reference category for the race-ethnicity variables. [c] For brevity, family and friend exposure is abbreviated to family exposure in this and subsequent tables.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

a second set of models, adding partner's exposures as predictors of mental health (e.g., mother exposure predicting youth PTSD symptoms). We report the results of these models as Model B in Table 5. By running the analyses in these two stages, we are able to isolate the impacts of own and partner exposure on PTSD and depression symptoms from each other while controlling for demographic characteristics.

Turning first to the prediction of youth mental health, we found that although youth's own direct exposure to 9/11 contributed significantly to the prediction of their self-reported depression symptoms, exposure did not predict youth's PTSD symptoms. Each of the minority race-ethnic groups reported more PTSD symptoms compared to European American youth, whereas youth who are second-generation immigrants reported fewer PTSD symptoms. Girls reported more, and first-generation immigrant youth fewer, depression symptoms. In neither of the two regressions for Model B was the F_{Δ} statistic significant; thus, none of mothers' exposures to 9/11 predicted youth PTSD or depression symptoms.

Mothers' own direct exposure to 9/11 did predict more of their own PTSD symptoms, over and above the contribution of demographic characteristics, $F(17, 321) = 5.04$, $p < .001$. Mothers' family and friend exposure and media exposure did not predict their own PTSD symptoms, and none of the forms of mothers' exposure to 9/11 predicted their depression symptoms. Mothers with higher levels of education reported fewer PTSD and depression symptoms, and employed mothers reported fewer depression symptoms; no other demographic characteristics predicted symptoms. The addition of youth exposure to the regressions did not improve any of the models significantly. Youth media exposure did predict mothers' PTSD symptoms (the more media exposure reported by youth, the fewer PTSD symptoms reported by mothers), although youth media exposure did not significantly improve the prediction of mothers' PTSD symptoms over and above demographic characteristics and mothers' own exposure.

Mothers' Coping Assistance as a Function of Exposure to 9/11

To determine whether mothers' coping assistance was in response to exposure to 9/11, we regressed each of the three forms of coping assistance first on mothers' exposure and then on youth's exposure. We followed a two-regression sequence similar to that used in the prediction of mental health outcomes, again to isolate the unique impacts of mothers' and youth's exposure to 9/11 on mothers' coping assistance. The results from these sets of regressions are presented in Table 6. Older mothers reported using emotional processing and roles and routines coping assistance less than younger mothers, whereas Hispanic American mothers reported using more distraction coping than European American mothers. Mother education was associated with less use of distraction coping, whereas mother employment was associated with more roles and routines coping.

The extent to which mothers engaged in emotional processing coping was significantly predicted from their own direct exposure to 9/11; the more direct exposure mothers experienced, the more they engaged in emotional processing coping with their adolescents ($\beta = .23$, $p < .001$). However, the extent to which mothers used roles and routines or distraction as methods of helping their children cope with the events of 9/11 was not predicted by the extent of their own exposure. Contrary to our prediction, mothers' coping assistance was not a function of the extent of youth's direct, family and friend, or media exposure to the events of 9/11; in other words, mothers did not engage in certain forms of coping assistance depending on the type of exposure their children experienced.

Mothers' Coping Assistance and Youth Mental Health

We examined whether mothers' coping assistance was associated with youth mental health by regressing youth PTSD symptoms and depression symptoms first on youth exposure to 9/11 and then on the three forms of mothers' coping. The pattern of significant demographic predictors was the same as that presented in Table 5. We again used two regressions to allow us to consider the possibility that mothers' coping assistance might function as mediators of the impact of youth exposure on their mental health. However, after the impact of youth's exposure to 9/11 on their mental health was controlled in Model A, none of the forms of mothers' coping assistance predicted youth mental health in either Model B (see Table 7).

Exposure to 9/11 and Youth Reports of Family and Personal Sociopolitical Attitudes

Youth's and mothers' exposure to 9/11 were used as predictors of youth reports of three

Table 6
Mothers' Ratings of Their Coping Assistance to Their Children Predicted From Their Own and Their Children's Exposure to 9/11

	Mothers' coping assistance					
	Emotional processing		Roles and routines		Distraction	
	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)
Adolescent's age	.07	.07	.02	.03	.04	.04
Adolescent's gender[a]	.01	−.01	.03	.01	.03	.03
Adolescent is immigrant	−.05	−.04	−.10	−.10	−.03	−.04
One or both parents are immigrants	.04	.02	−.04	−.05	−.02	−.01
Is African American[b]	−.05	−.00	.01	.04	.12	.11
Is Hispanic American[b]	−.10	−.05	.02	.05	.30**	.29**
Is of Another race or ethnicity[b]	.02	.04	.01	.03	.07	.07
Mother's age	−.11*	−.11*	−.13*	−.13*	−.04	−.04
Mother is married	−.03	−.02	.02	.02	.05	.05
Mother's level of education	−.10	−.09	−.02	−.02	−.15*	−.15*
Mother is employed	.08	.08	.12*	.12*	.10	.10
Time between 9/11 and the interview[b]	−.06	−.07	−.04	−.05	.05	.05
RCCP intervention—Lessons	−.02	−.02	.04	.04	−.05	−.05
RCCP intervention—Staff development	−.04	−.04	−.03	−.03	.02	.02
Mothers' exposure to 9/11						
Direct exposure	.23***	.22***	.10	.09	.08	.08
Family exposure	.00	−.01	.07	.06	−.02	−.02
Media exposure	.07	.05	.07	.07	.09	.09
Youth's exposure to 9/11						
Direct exposure	—	.05	—	.05	—	.02
Family exposure	—	.08	—	.04	—	−.03
Media exposure	—	.06	—	−.01	—	−.02
R^2	.10	.12	.07	.08	.11	.11
F	2.44***	2.35***	1.68*	1.49	2.68***	2.28***
df	17, 358	20, 355	17, 358	20, 355	17, 358	20, 355
R^2_Δ	—	.01	—	.00	—	.00
F_Δ	—	1.75	—	0.49	—	0.14
df_Δ	—	3	—	3	—	3

Note. RCCP = Resolving Conflict Creatively Program.
[a] 0 = *male*, 1 = *female*. [b] These characteristics were for mothers, with European Americans as the reference category for the race-ethnicity variables.
*$p < .05$. **$p < .01$. ***$p < .001$.

aspects of family sociopolitical attitudes, namely family lessons about prejudice, family discussions about current events, and family discussions about social mistrust. Girls, African Americans, and Hispanic Americans were most likely to report their parents taught them about prejudice and about needing to mistrust others. Youth whose parents are immigrants report more trust in their local communities, whereas both African American and Hispanic American youth reported less trust. Youth whose mothers were married or working reported less prejudice toward immigrants. Interestingly, youth's prejudice toward immigrants was higher the longer the gap between 9/11 and their interviews.

The more youth reported media exposure to 9/11, the more likely they were to agree that their family discussed circumstances in which they might personally encounter prejudice, as well as the extent to which they should be vigilant about

Table 7

Youth's PTSD and Depression Symptoms Predicted From Their Own Exposure to 9/11 and Mothers' Coping Assistance

	Youth's PTSD symptoms		Youth's depression symptoms	
	Model A: Own exposure (β)	Model B: Adding mothers' coping assistance (β)	Model A: Own exposure (β)	Model B: Adding mothers' coping assistance (β)
Adolescent's age	.07	.08	.06	.05
Adolescent's gender[a]	.07	.07	.16**	.16**
Adolescent is immigrant	−.10	−.09	−.17**	−.17**
One or both parents are immigrants	−.13*	−.13*	−.05	−.05
Is African American[b]	.25**	.25**	−.04	−.04
Is Hispanic American[b]	.30**	.31**	−.02	−.01
Is of another race or ethnicity[b]	.15*	.16*	−.02	−.02
Mother's age	.08	.08	.03	.03
Mother is married	−.02	−.02	.03	.03
Mother's level of education	−.04	−.05	−.07	−.07
Mother is employed	.08	.08	.03	.04
Time between 9/11 and the interview[b]	−.06	−.06	−.00	−.00
RCCP intervention—Lessons	.04	.04	.06	.06
RCCP intervention—Staff development	−.01	−.00	.01	.01
Youth's exposure to 9/11				
Direct exposure	.09	.09	.12*	.12*
Family exposure	.02	.02	−.02	−.02
Media exposure	.04	.04	.03	.03
Mothers' coping assistance				
Emotional processing	—	−.00	—	.06
Roles and routines	—	.03	—	−.08
Distraction	—	−.05	—	.00
R^2	.08	.08	.08	.09
F	1.79*	1.54	1.86*	1.67*
df	17, 360	20, 357	17, 360	20, 357
R^2_Δ	—	.00	—	.01
F_Δ	—	0.20	—	0.60
df_Δ	—	3	—	3

Note. PTSD = posttraumatic stress disorder; RCCP = Resolving Conflict Creatively Program.

[a] 0 = *male*, 1 = *female*. [b] These characteristics were for youth, with European Americans as the reference category for the race-ethnicity variables.

$*p < .05$. $**p < .01$. $***p < .001$.

and distrustful of other people (see Table 8). Youth media exposure was also associated with more family discussions of current events. Mothers' exposure to 9/11 was not predictive of any of the youth-reported measures of family sociopolitical attitudes. Regarding youth's self-reported sociopolitical attitudes, youth's media exposure significantly predicted youth's trust in the local community. The only sociopolitical attitude affected by youth's direct exposure to 9/11 was their prejudice toward immigrants, although it did not account for a significant improvement in the R^2. Contrary to expectation, more exposed youth reported less, not more, prejudice toward immigrants.

Exposure to 9/11 and Mother Reports of Family and Personal Sociopolitical Attitudes

Similar to the models predicting youth reports, mothers' reports of the extent of their own socialization of their adolescents around issues of prejudice, social mistrust, and current events were predicted significantly by their own media exposure to 9/11 (see Table 9). In addition, mothers' family and friend exposure predicted a greater extent of mother-reported family lessons about prejudice, whereas the more mothers reported direct exposure to 9/11, the less likely they were to discuss issues of social mistrust with their adolescents. Youth exposure did not significantly

Table 8

Youth's Reports of Family and Personal Sociopolitical Attitudes Predicted From Youth's and Mothers' Exposure to 9/11

	Youth's reports of family and personal sociopolitical attitudes									
	Family lessons about prejudice		Family lessons about social mistrust		Family discussions about current events		Trust in local community		Prejudice toward immigrants	
	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)	Model A: Own exposure (β)	Model B: Adding mothers' exposure (β)
Adolescent's age	.08	.08	.06	.05	.11*	.10	–.07	–.08	–.01	–.00
Adolescent's gender[a]	.10*	.10	.17***	.17***	.05	.05	–.07	–.06	–.02	–.02
Adolescent is immigrant	–.04	–.04	–.04	–.04	–.07	–.08	.04	.04	–.04	–.03
One or both parents are immigrants	–.05	–.05	–.04	–.03	–.06	–.05	.14*	.15*	–.06	–.07
Is African American[b]	.25**	.25**	.20*	.19*	–.02	–.03	–.18*	–.19*	–.09	–.07
Is Hispanic American[b]	.24*	.23*	.23*	.22*	–.02	–.03	–.25*	–.25*	–.14	–.14
Is of another race or ethnicity[b]	.09	.09	.02	.02	.04	.04	–.06	–.05	–.09	–.08
Mother's age	–.02	–.02	–.06	–.06	–.00	–.01	.02	.02	.01	.01
Mother is married	–.05	–.04	–.03	–.03	.09	.09	–.02	–.02	–.12*	–.13*
Mother's level of education	.00	.00	–.03	–.03	–.04	–.04	.06	.05	–.12*	–.13*
Mother is employed	–.11	–.11*	–.10	–.11	–.09	–.09	–.04	–.05	.06	–.07
Time between 9/11 and the interview[b]	–.11	–.10	–.09	–.08	.01	.01	–.03	–.02	.15**	.13*
RCCP intervention—Lessons	–.03	–.03	–.03	–.03	–.09	–.10	–.18*	–.20**	.10	.09
RCCP intervention—Staff development	.02	.02	.00	.02	.04	.05	.15*	.17*	–.04	–.04
Youth's exposure to 9/11										
Direct exposure	.10	.10	.02	.02	.08	.08	.07	.06	–.11*	–.10
Family exposure	.09	.09	.05	.05	.07	.06	.09	.06	.05	.04
Media exposure	.15**	.15**	.15**	.14**	.24***	.23***	.11*	.10	–.01	–.01
Mothers' exposure to 9/11										
Direct exposure	—	.02	—	.02	—	.04	—	.06	—	–.11
Family exposure	—	–.02	—	.01	—	–.01	—	.08	—	.08
Media exposure	—	–.01	—	.06	—	.05	—	.08	—	.03
R^2	.12	.12	.12	.12	.12	.13	.10	.12	.08	.09
F	2.76***	2.34***	2.84***	2.48***	3.02***	2.62***	2.33***	2.42***	1.92*	1.85*
df	17, 363	20, 360	17, 363	20, 360	17, 363	20, 360	17, 363	20, 360	17, 363	20, 360
R^2_Δ		.00		.00		.00		.02		.01
F_Δ		0.08		0.52		0.47		2.73*		1.44
df_Δ		3		3		3		3		3

Note. RCCP = Resolving Conflict Creatively Program.

[a] 0 = *male*, 1 = *female*. [b] These characteristics were for youth, with European Americans as the reference category for the race-ethnicity variables.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 9

Mothers' Reports of Family and Personal Sociopolitical Attitudes Predicted From Mothers' and Youth's Exposure to 9/11

	Mothers' reports of family and personal sociopolitical attitudes									
	Family lessons about prejudice		Family lessons about social mistrust		Family discussions about current events		Trust in local community		Prejudice toward immigrants	
	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)	Model A: Own exposure (β)	Model B: Adding youth's exposure (β)
Adolescent's age	−.03	−.03	.02	.02	−.09	−.08	.04	.04	−.04	−.04
Adolescent's gender[a]	−.06	−.07	−.01	−.01	−.05	−.05	−.09	−.08	−.08	−.09
Adolescent is immigrant	.02	.02	.02	.03	−.03	−.03	.04	.04	.08	.08
One or both parents are immigrants	−.08	−.09	−.03	−.03	.01	.00	.13*	.14*	.21***	.21***
Is African American[b]	.14	.17*	.22**	.24**	.09	.09	−.29***	−.31***	−.09	−.08
Is Hispanic American[b]	.08	.12	.28**	.30**	.02	.01	−.33***	−.35***	−.00	.01
Is of another race or ethnicity[b]	−.06	−.04	.11	.12*	−.00	−.01	−.12*	−.13*	−.05	−.04
Mother's age	.05	.05	−.06	−.06	.05	.05	.13*	.13*	.06	.06
Mother is married	−.03	−.03	−.11*	−.11*	.01	.01	.01	.01	.00	.00
Mother's level of education	.02	.02	−.04	−.04	.20***	.20***	.03	.03	.08	.08
Mother is employed	.04	.05	−.00	.01	−.03	−.04	.07	.07	.05	.04
Time between 9/11 and the interview[b]	.06	.07	.08	.08	−.03	−.03	−.11*	−.11	.05	.05
RCCP intervention—Lessons	.03	.03	−.08	−.07	−.15*	−.15*	−.05	−.04	.04	.03
RCCP intervention—Staff development	.08	.08	.14	.13	.07	.06	.08	.07	−.04	−.03
Mothers' exposure to 9/11										
Direct exposure	−.03	−.03	−.12*	−.11*	.05	.05	.07	.08	.06	.04
Family exposure	.13*	.12*	.05	.03	.04	.05	.07	.07	.01	.02
Media exposure	.12*	.13*	.19***	.19***	.28***	.26***	−.02	−.02	−.11*	−.11*
Youth's exposure to 9/11										
Direct exposure	—	.02	—	−.07	—	−.06	—	−.07	—	.07
Family exposure	—	.09	—	.10	—	.02	—	.00	—	−.02
Media exposure	—	−.03	—	.02	—	.09	—	.00	—	.03
R^2	.08	.09	.14	.15	.18	.19	.13	.13	.09	.10
F	1.88*	1.75*	3.34***	3.09***	4.66***	4.15***	3.18***	2.79***	2.17**	1.96**
df	17, 362	20, 359	17, 362	20, 359	17, 362	20, 359	17, 632	20, 359	17, 632	20, 359
R^2_Δ	—	.01	—	.01	—	.01	—	.00	—	.01
F_Δ	—	1.01	—	1.59	—	1.21	—	0.63	—	0.77
df_Δ	—	3	—	3	—	3	—	3	—	3

Note. RCCP = Resolving Conflict Creatively Program.
[a]0 = *male*, 1 = *female*. [b]These characteristics were for mothers, with European Americans as the reference category for the race-ethnicity variables.
*$p < .05$. **$p < .01$. ***$p < .001$.

improve the prediction when added in the second models.

In the models predicting mothers' own sociopolitical attitudes, neither mothers' own nor youth's exposure to 9/11 predicted mothers' trust in the local community. Mothers' prejudice toward immigrants was predicted by their media exposure, with more exposure associated with less prejudice toward immigrants. Youth exposure to 9/11 did not predict mothers' self-reported trust in the local community or mothers' self-reported prejudice toward immigrants.

African American and Hispanic American mothers reported higher levels of talking with their children about social mistrust and less likely to have trust in their local community. Mothers who are immigrants or married to an immigrant report more trust in their local communities and, somewhat surprisingly, more prejudice toward fellow immigrants. Older mothers report more trust in their communities, married mothers report fewer family lessons about social mistrust, and educated mothers report more family discussions about current events.

Discussion

The 9/11 attacks live on in the memories of many Americans and most New Yorkers. Our main conclusion from this study is thus a surprising one—namely that variations in exposure to the events of 9/11 had very modest associations with the longer term mental health of New York youth and their mothers. We also found that variation in exposure to the events of 9/11 was associated to a small degree with both mothers' coping assistance and youth's and mothers' sociopolitical attitudes. We found little evidence of reciprocal influences of mothers' or youth's exposure on each others' mental health and sociopolitical attitudes.

Exposure to 9/11 and Parent and Youth Mental Health

Although other studies have found that media exposure alone to the 9/11 attacks induced PTSD symptoms in children geographically distant from the sites of the attacks (Otto et al., 2007), we did not replicate this finding with our NYC sample, instead determining that only direct exposure was associated with youth's and mothers' mental health. It may be that direct exposure to the attacks overshadowed the impacts of media exposure among individuals so proximal to the attack site. The physical reminders of the attack, including the smoke rising from what was left of the towers, the restricted access to Lower Manhattan, the rerouted subways, and the pictures of missing victims of the attacks posted around the city in vain by family members, were present for days and weeks afterward, prolonging exposure to the attacks among New Yorkers. It may also be that families living so close to the site of the attacks felt less reason than Americans living elsewhere to turn to the media to learn about the attack, as they need only go to any street corner, subway car, diner, or office lunchroom to hear the latest information on victim counts, conspiracy theories, and warnings of future attacks.

Despite nearly universal direct exposure to the 9/11 attacks, very little of the variance in the mental health symptoms of youth and parents in our study was explained by 9/11 exposure. Even for the two models in which exposure to 9/11 significantly predicted symptomatology, exposure accounted for only 2% of the variance in youth depression symptoms and 7% of the variance in mothers' PTSD symptoms. The lack of association between 9/11 exposure and mental health symptoms may be a function of the length of time between 9/11 and the interviews (on average 15 months for both youth and their mothers). Although we controlled for the time lapse between 9/11 and each interview, we identified few enduring associations of 9/11 exposure with youth's and mothers' mental health. It is still quite possible, if not likely, that there were more immediate effects of 9/11 exposure on the mental health of our participants, but our results suggest that any such impacts on PTSD and depression symptoms have largely receded with the passage of time.

Given the nontrivial rates of PTSD (youth: 4%; mothers: 20%) and depression (youth: 12%; mothers: 34%) symptoms in this sample of youth and their mothers living in NYC, it is clear that something other than the trauma of the 9/11 attacks is responsible for these elevated levels of problematic mental health symptoms. In our previous empirical paper from this study (Aber, Gershoff, Ware, & Kotler, 2004), we explored community violence exposure as a potential predictor and found that exposure to community violence as a victim or witness significantly predicted PTSD and depression symptoms, with near zero associations for 9/11 exposure dwarfed by the very large effects of violence victimization. We concluded then that the effects of traumatic events such as disasters and terrorist attacks need to be understood in the context of youth's everyday lives. Similarly, parents' PTSD and depression symptoms may be predicted more by daily exposures to stress or violence in their homes and communities. Unfortunately, we did not have measures of mothers' individual violence exposure in this study and so could not test this alternative hypothesis; such a question is ripe for future research. We encourage future researchers of the impacts of disasters to include measures of "everyday" stress and violence to put the impacts of disasters in context.

Other studies of 9/11 have found similarly modest associations of 9/11 exposure with child and youth mental health. One longitudinal study of children in NYC found that parents reported fewer child behavior problems 4 months after the attack, but by 6 months later behavior problems had returned to pre-9/11 levels (Stuber et al., 2005). Similarly, a study of high school students in Boston found that while political interest and civic participation increased slightly right after 9/11, they returned to pre-9/11 levels by 9 months after the attacks (Metz & Youniss, 2003). A study of 9- to 13-year-olds similarly found that neither knowing someone directly affected by the attacks nor having moderate media exposure to the 9/11 attacks predicted children's PTSD symptoms 7 months after 9/11 (Lengua, Long, & Meltzoff, 2006). In that study, preattack stress load (e.g., moving, serious illness/injury, death of a loved one) and the extent of children's 9/11-specific threat appraisal and avoidant coping were much stronger predictors of PTSD symptoms. Among fourth graders living in inner-city Chicago at the time of the 9/11 attack, those assessed post-9/11 did not have higher levels of depression symptoms than the children assessed pre-9/11 (Henry et al., 2004).

The good news of our null findings is that even among a sample of families with extensive direct exposure to the 9/11 terrorist attacks and its aftermath, there were few lasting associations of this exposure with mental health on average 1 year later. We may be tempted to conclude that youth and their mothers are "resilient" to the impacts of 9/11. However, resilience is a dynamic construct that involves positive adaptation after an experience of adversity (Luthar, Cicchetti, & Becker, 2000). To determine whether our participants are in fact doing as well as they were before 9/11, we would need pre- and postmeasures of functioning. Without preevent measures, we are only able to predict post-9/11 levels of functioning and not change or adaptation of functioning after 9/11. We further argue that any conclusion about resilience in our sample is belied by the fact that the mothers and youth in our sample reported high levels of symptomatology over 1 year after the 9/11 attack. If 9/11 is not the cause of these mental health problems, what is? We suggest that greater attention needs to be paid to everyday stress and trauma in the lives of inner-city families.

Exposure to 9/11 and Parent Coping Assistance

Mothers of youth in New York did report small differences in their modes of helping their children cope with 9/11 depending on their own exposure. Specifically, mothers' higher direct exposure was associated with higher levels of coping focused on emotions. Yet in contrast to the literature that describes an important role for parents' coping styles in helping children and youth adjust after a disaster (Compas et al., 1992; Vernberg et al., 1996), we found that mothers' self-reported coping assistance to their adolescents was not associated with adolescents' PTSD and depression symptoms. Ours is not the only study to not find positive outcomes associated with parental coping; in a study of adolescents whose families were displaced by Hurricane Katrina, only one of five forms of parental coping was associated with youth mental health, namely the extent to which parents sought professional help for their adolescents, and the association was with higher distress and depression scores (Vigil & Geary, 2008).

Exposure to 9/11 and Sociopolitical Attitudes

In contrast to the meager associations between exposure to the events of 9/11 and mental health, there were more and stronger associations between youth's and mothers' exposure and their family and personal sociopolitical attitudes. Of the three types of youth's exposure, media exposure had the clearest connection to youth attitudes, being associated with four of the five sociopolitical constructs measured, namely family lessons about prejudice, mistrust, and current events as well as youth's trust in their local communities. Youth with more media exposure had families more actively engaged in a political socialization process after the events of 9/11. Youth family and friend exposure and direct exposure were, in large part, not associated with families' sociopolitical socialization or with youth's sociopolitical attitudes.

A notable exception to this pattern was the association between youth direct exposure and less prejudice toward immigrants. This finding suggests that public concern that a consequence of 9/11 would be an increase in xenophobia was misplaced. Interestingly, youth's prejudice toward immigrants was not predicted by their own or their parents' immigrant status; given that fully 64% of the youth had at least one parent who was born outside the United States and thus are themselves second-generation immigrant youth, it may be that generational status does not differentiate among these youth. It is also important to note that we asked about prejudice toward immigrants in general, not toward any specific immigrant groups. If we had asked specifically about prejudice toward Muslims or Arabs, the groups most associated with the 9/11 terrorist attackers, we may very well have found that exposure to 9/11 predicted more prejudice toward these groups.

Mothers' reports of family and personal sociopolitical attitudes were associated with their own exposure. As with their adolescents, mothers' media exposure was most strongly and consistently predictive of the extent to which mothers report they socialize their children regarding the issues of prejudice, social mistrust, and current events. For mothers, media exposure predicted less prejudice toward immigrants, and direct exposure predicted less socialization around social mistrust.

Taken together, this pattern of findings suggests that, while the events of 9/11 were associated with limited enduring associations with the mental health of youth and their mothers, 9/11 exposure did predict mothers' socialization of and youth's development of sociopolitical attitudes. It is important to note, however, that any significant associations with 9/11 exposure occurred largely within persons (in the form of individual effects) rather than across persons (as reciprocal effects).

Limitations

Several aspects of the study limit the strength of our conclusions. First, all of the data were collected via self-reports and so are susceptible to reporter bias. For instance, if more sociopolitically engaged mothers and adolescents are more media attentive, the association between media exposure and sociopolitical attitudes could be epiphenomenal. Second, while the events of 9/11 constitute a truly uncontrollable, exogenous shock, recalling exposure is not purely exogenous. The associations we report are vulnerable to selection effects. Although we would have preferred to model selection into our models, data limitations prevented us from doing so. However, by reporting the extent to which individual characteristics predicted "selection" into 9/11 exposure and by including these characteristics in our models, we have accounted for selection effects.

There are also many limitations characteristic of disaster studies, particularly inadequate predisaster "baseline" data (see Aber & Gershoff, 2004; Gershoff & Aber, 2004, for more comprehensive review of such limitations), which this study was unable to transcend. As noted earlier, context matters in understanding the effect of systemic violence and man-made disaster on child and parent development (Aber et al., 2004). This was not a study of the distant trauma of 9/11 but rather was a study of parents and youth who were in NYC on 9/11 and thus in close proximity to the attacks on the WTC. Nor was this a study of a demographically representative sample of NYC children (see Hoven et al., 2005, for such a study). Rather, this was a study of primarily Hispanic American and African American families from low-income communities. These contextual factors limit our ability to generalize these findings to broader populations. But they also help generate ideas about how contextual factors may moderate some of the influences we investigated in this study. To help inform not only developmental theory and research but also program and policy initiatives related to children and parents in disasters or other emergencies, the field needs to develop a more complete and unifying theoretical frame for the integration of such studies as Norris and colleagues (Norris, Friedman, Watson, Byrne, et al., 2002) have begun to do.

Conclusion

We have used rich data on three forms of exposure to the 9/11 terrorist attacks to examine whether such exposure had long-term impacts on youth and their families living, working, and going to school in relatively close proximity to the site of the WTC attacks. Rather than finding evidence of enduring impairment following from the attacks on 9/11, we have found a silver lining. Exposure to the 9/11 attacks appears to have prompted these families to have more discussions about serious social and political issues, discussions that we hope will lead to a more informed and civically minded generation of young adults.

References

Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology, 29*, 324–348.

Aber, J. L., & Gershoff, E. T. (2004). Editors' conclusion: Child, youth, and parent responses to the terrorism of September 11th 2001: Implications for applied developmental science and practice. *Applied Developmental Science, 8*, 221–225.

Aber, J. L., Gershoff, E. T., Ware, A., & Kotler, J. A. (2004). Estimating the effects of September 11th and other forms of violence on the mental health and social development of New York City's youth: A matter of context. *Applied Developmental Science, 8*, 111–129.

Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Developmental Psychopathology, 10*, 187–213.

Brown, J. L., Roderick, T., Lantieri, L., & Aber, J. L. (2004). The Resolving Conflict Creatively Program: A school-based social and emotional learning program. In J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.), *Building academic success on social and emotional learning: What does the research say?* (pp. 151–169). New York: Teachers College Press.

C-DISC Development Group. (2000). *Diagnostic interview schedule for children: Psychometrics*. New York: NIMH-DISC Training Center at Columbia University/New York State Psychiatric Institute.

Compas, B. E., Worsham, N. L., & Ey, S. (1992). Conceptual and developmental issues in children's coping with stress. In A. M. La Greca, L. J. Siegel, J. Wallander, & C. E. Walker (Eds.), *Stress and coping in child health* (pp. 7–24). New York: Guilford.

Federal Emergency Management Agency. (2004, August). *Helping children cope with disaster*. Retrieved June 18, 2008, from http://www.fema.gov/pdf/library/children.pdf

Flanagan, C., Gill, S., & Gallay, L. (1998, November). *Intergroup understanding, social justice, and the "social contract" in diverse communities of youth: Foundations for civic understanding*. Project report prepared for the Workshop on Research to Improve Intergroup Relations Among Youth, Forum on Adolescence, Board on Children, Youth, and Families, National Research Council, Washington, DC.

Flanagan, C., Ingram, P., Gallay, E. M., & Gallay, E. E. (1997). Why are people poor? Social conditions and adolescents' interpretation of the "social contract." In R. D. Taylor & M. C. Wang (Eds.), *Social and emotional adjustment and family relations in ethnic minority families* (pp. 53–62). Mahwah, NJ: Erlbaum.

Galea, S., Ahern, J., Resnick, J., Kilpatrick, D., Bucuvalas, M., Gold, J., et al. (2002). Psychological sequelea of the September 11 terrorist attacks in New York City. *New England Journal of Medicine, 346*, 982–987.

Gershoff, E. T., & Aber, J. L. (2004). Editors' introduction: Assessing the impact of September 11th, 2001, on children, youth and families: Methodological challenges to research on terrorism and other non-normative events. *Applied Developmental Science, 8*, 106–110.

Gould, M. S., Munfakh, J. L. H., Kleinman, M., Lubell, K., & Provenzano, D. (2004). Impact of the September 11th terrorist attacks on teenagers' mental health. *Applied Developmental Science, 8*, 158–169.

Hendricks, C., & Bornstein, M. H. (2007). Ecological analysis of early adolescents' stress responses to 9/11 in the Washington, DC, area. *Applied Developmental Science, 11*, 71–88.

Henry, D. B., Tolan, P. H., & Gorman-Smith, D. (2004). Have there been lasting effects associated with the September 11, 2001, terrorist attacks among inner-city parents and children? *Professional Psychology, Research, and Practice, 35*, 542–547.

Hoven, C. W., Duarte, C. S., Lucas, C. P., Wu, P., Mandell, D. J., Goodwin, R. D., et al. (2005). Psychopathology among New York City public school children six months after September 11. *Archives of General Psychiatry, 62*, 545–552.

Hoven, C. W., Duarte, C. S., Mandell, D., Musa, G., Wicks, J., Wu, P., et al. (2002). *WTC-NYC Child and Adolescent Questionnaire*. Unpublished measure. New York: Columbia University.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (2002). Introduction. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 11–33). Washington, DC: American Psychological Association.

Lengua, L. J., Long, A. C., & Meltzoff, A. N. (2006). Pre-attack stress-load, appraisals, and coping in children's responses to the 9/11 terrorist attacks. *Journal of Child Psychology and Psychiatry, 47*, 1219–1227.

Lucas, C. P., Zhang, H., Fisher, P., Shaffer, D., Regier, D., Narrow, W., et al. (2001). The DISC Predictive Scales (DPS): Efficiently predicting diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 443–449.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543–562.

Metz, E., & Youniss, J. (2003). September 11 and service: A longitudinal study of high school students' views and responses. *Applied Developmental Science, 7*, 148–155.

Norris, F. H., Friedman, M. J., & Watson, P. J. (2002). 60,000 disaster victims speak: Part II. Summary and implications of the disaster for mental health research. *Psychiatry, 65*, 240–260.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. Z. (2002). 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981-2001. *Psychiatry, 65,* 207–239.

Otto, M. W., Henin, A., Hirshfeld-Becker, D. R., Pollack, M. H., Biederman, J., & Rosenbaum, J. F. (2007). Posttraumatic stress disorder symptoms following media exposure to tragic events: Impact of 9/11 on children at risk for anxiety disorders. *Journal of Anxiety Disorders, 21,* 888–902.

Phillips, D., Featherman, D. L., & Liu, J. (2004). Children as an evocative influence on adults' reactions to terrorism. *Applied Developmental Science, 8,* 195–210.

Phillips, D., Prince, S., & Schiebelhut, L. (2004). Elementary school children's responses 3 months after the September 11 terrorist attacks: A study in Washington, DC. *American Journal of Orthopsychiatry, 74,* 509–528.

Prinstein, M. J., La Greca, A. M., Vernberg, E. M., & Silverman, W. K. (1996). Children's coping assistance: How parents, teachers, and friends help children cope after a natural disaster. *Journal of Clinical Child Psychology, 25,* 463–475.

Putnam, R. D. (2002, February 11). Bowling together. *The American Prospect, 13,* 20–22.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Rasinski, K. A., Berktold, J., Smith, T. W., & Albertson, B. L. (2002, August 7). *America recovers: A follow-up to a national study of public response to the September 11th terrorist attacks.* Chicago: NORC, University of Chicago.

Schlenger, W. E., Caddell, J. M., Ebert, L., Jordan, K., Rourke, K. M., Wilson, D., et al. (2002). Psychological reactions to terrorist attacks: Findings from the National Study of Americans' Reactions to September 11. *Journal of the American Medical Association, 288,* 581–588.

Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab-Stone, M. (2000). NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 28–38.

Silver, R. C., Holman, A., McIntosh, D. N., Poulin, M., & Gil-Rivas, V. (2002). Nationwide longitudinal study of psychological responses to September 11. *Journal of the American Medical Association, 288,* 1235–1244.

Stuber, J., Galea, S., Pfefferbaum, B., Vandivere, S., Moore, K., & Fairbrother, G. (2005). Behavior problems in New York City's children after the September 11, 2001, terrorist attacks. *American Journal of Orthopsychiatry, 75,* 190–200.

Traugott, M., Brader, T., Coral, D., Curtin, R., Featherman, D., Groves, R., et al. (2002, September). How Americans responded: A study of public reactions to 9/11/01. *Political Science and Politics, 35,* 511–516.

Vernberg, E. M., La Greca, A. M., Silverman, W. K., & Prinstein, M. J. (1996). Prediction of posttraumatic stress symptoms in children after Hurricane Andrew. *Journal of Abnormal Psychology, 105,* 237–248.

Vigil, J. M., & Geary, D. C. (2008). A preliminary investigation of family coping styles and psychological well-being among adolescent survivors of Hurricane Katrina. *Journal of Family Psychology, 22,* 176–180.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1161–1175

The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories

Elif Celebi Oncu
Oregon State University

Aysegul Metindogan Wise
Bogazici University

The purpose of this exploratory study was to determine whether projective techniques could identify long-term consequences among children stemming from exposure to a traumatic event. The first group of children ($n = 53$; 26 female, 27 male) experienced 2 major earthquakes at age 7, 3 months apart, in Turkey, while a similarly matched control group ($n = 50$; 25 female, 25 male) did not. Both groups of children (current age: 9) completed a series of short stories related to disastrous events. Results indicated that the traumatized group evinced a range of trauma-related symptoms 2 years after experiencing the earthquakes.

On August 17, a 7.4-magnitude earthquake occurred in Golcuk, Kocaeli, with effects felt in the neighboring provinces of Yalova, Istanbul, Sakarya, Bursa, Bolu, and Eskisehir. Soon after the Golcuk earthquake, on November 12, a second earthquake hit the neighboring province of Duzce with a magnitude of 7.2. The high magnitudes of the earthquakes along with the poor infrastructure in the affected areas resulted in tremendous losses. According to the United Nations Childrens Fund (UNICEF; 2001), 18,000 people died, 44,000 were injured, and more than 600,000 people were left homeless out of a total of 2 million people. The effects of the earthquake were exacerbated when the roads and railways became unusable, water and sewage were cut, and the electricity grid no longer functioned. According to the Earthquake Research Department of the General Directorate of Disaster Affairs, the financial damage caused by natural disasters, including earthquakes, in Turkey within the last 10 years is US$20 billion (http://www.deprem.gov.tr/dogalafet.htm). It is estimated that US$1.5 billion of that loss is due to the damage that was caused to the infrastructure of the earthquake-hit areas (UNICEF, 2001).

Because the North Anatolian fault line stretches across Turkey, geologists expect there will be an even bigger earthquake (referred to as the ''Marmara earthquake'') to occur in the northwestern side of the country within the next 30 years, including both the previous earthquake areas and Istanbul, a city of 15 million people. Thus, especially recently, researchers have been focusing on identifying the effects of these earthquakes and the ways in which adults and children deal with them (Basoglu, Salcioglu, & Livanou, 2002; Bulut, Bulut, & Tayli, 2005; Kilic et al., 2006). Kilic et al. (2006) focused on adults and found that relocation after the earthquake was associated with depression and posttraumatic stress disorder (PTSD). On the other hand, Bulut et al. (2005) found that children experienced PTSD even though there was no relocation. It appears that adults and children may be affected by the disaster differently and it is important to further investigate the delayed effects of earthquake in order to be prepared for a possible future disaster and to deal with its consequences.

The Emotional Consequences of Trauma or Disaster

Although traumatic events differ in nature, based on whether they are human-made or natural, the effects of traumatic events show some similarities in the effects they produce. One of the reasons natural disasters affect children more dramatically than many human-made trauma, such as abuse or neglect, is that natural disasters such as earthquakes occur fast and unexpectedly (Durkin, Khan, Davidson, Zaman, & Stein, 1993; Monahon, 1997). Moreover, these unexpected events affect people's moods directly, especially children's, and can cause

---

Correspondence concerning this article should be addressed to Elif Celebi Oncu, University of Kocaeli, Egitim Fak, Ilkogretim Bol, Umuttepe Merkez Yerleskesi, 41380 Izmit, Kocaeli, Turkey. Electronic mail may be sent to ecelebioncu@yahoo.com.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0010

trauma (Azarian & Skriptchenko-Gregorian, 1998; Durkin et al., 1993; Jones, Frary, Cunningham, Weddle, & Kaiser, 2001; Mercuri & Angelique, 2004; Monahon, 1997; Yule, 2001). The extent and severity with which natural disasters affect children depend on numerous factors, including the child's age, gender (Gray Deering, 2000; Monahon, 1997; Vizek-Vidovic, Kuterovac-Jagodic, & Arambasic, 2000), developmental level (Falasca & Caulfield, 1999), education, and personality. Previous studies showed that boys have been found to react with more aggression and externalizing problem behaviors whereas girls are reported to respond to traumatic events with internalizing problem behaviors such as depression (Monahon, 1997; Vizek-Vidovic et al., 2000; Winje & Ulvik, 1998).

Some studies investigating children's reactions to disasters have demonstrated that the degree of emotional reaction to trauma is related to the proximity of the disaster (Jones et al., 2001; Najarian, Goenjian, Pelcovitz, Mandel, & Najarian, 1996). However, other studies indicate that it is not proximity to the disaster but rather the subjective experience of the disaster that predicts PTSD (Gokben Hizli, Taskintuna, Isikli, Kilic & Zileli, 2008). Gokben Hizli et al. (2008) found that depressive symptoms of adolescents endured 4 years after the earthquake in Turkey as a function of subjective negative perceptions of the disaster. Neria, Nandi, and Galea (2008) conducted a systematic review of studies that focused on investigating posttraumatic stress after experiencing a disaster between 1980 and 2007. Of 284 studies, there were three different identified categories of disasters: human-made, technological, and natural disasters. Neria et al. (2008) argued that although there seems to be some differences in the outcome, the effects of natural disasters were similar to the effects of human-made and technological disasters. For example, it appears that PTSD is less common after a natural disaster compared to the other types of disasters. However, the authors argue, when an earthquake hits, it is a larger scale disaster than human-made and technological disasters; thus, studies include victims who are in the center and experience the disaster directly and those who are further away from the center and experienced it indirectly. As a consequence, it seems that whether a negative event is experienced as a disaster is related to how many people it affects and how serious the consequences of the negative event are rather than whether the original cause was human-made, technological, or natural (Neria et al., 2008). In a study conducted by Mercuri and Angelique (2004), the PTSD levels of children were examined among natural, technological, and nontechnological disasters. It was found that children had exhibited increased levels of PTSD when they witnessed scenes of destruction and/or experienced life threat regardless of the type of disaster. And they found in the studies that it is important to ask children directly about their postdisaster reactions because none of the studies they had examined, in which parent measures were used, reported any significant PTSD in children.

Previous disaster research shows that traumatized children displayed problems of emotional and behavioral disturbances such as flaring tempers phobias, anxiety, sleep and eating disorders, and enuresis (Azarian & Skriptchenko-Gregorian, 1998; Durkin et al., 1993; Goenjian et al., 1995; Gurwitch, Silovsky, Shultz, Kees, & Burlingame, 2002; McAdam Crisp, 2006; Shen & Sink, 2002; Yule, 2001). Although most children are affected by the consequences of natural disasters, effects of disasters and perceptions of traumatizing events can vary from child to child. Factors such as age, gender, education, and personality can determine the degree of the trauma experienced by children (Giannopoulou et al., 2006). Boys have been found to react with more aggression and externalizing problem behaviors whereas girls are reported to respond to traumatic events with more internalizing problem behaviors such as depression (Giannopoulou et al., 2006; Monahon, 1997; Vizek-Vidovic et al., 2000; Winje & Ulvik, 1998).

Children's reactions to a disaster vary in quality depending on whether they are preschoolers, school-aged children, or adolescents (Vizek-Vidovic et al., 2000). Gray Deering (2000) suggested that preschool children had difficulty understanding disasters. In addition, Gray Deering (2000) argued that preschool children are egocentric in their thinking and as a result they can feel guilty and blame themselves for what happened. Young children's reactions to disaster can also include antisocial behaviors, aggression, and regression (Gray Deering, 2000; Yule, 2001). School-aged children have shown aggression, withdrawal, feelings of anxiety, fear, helplessness, grief, and worries about safety of self and others, and worries about reoccurrence of the traumatic event (Durkin et al., 1993; Pynoos, Goenjian, & Steinberg, 1998; Shaw, Applagate, & Schorr, 1996; Shen & Sink, 2002). Thus, children differ in their reactions to trauma and in their behaviors after the trauma is experienced based on their age and their level of understanding of the negative event (Giannopoulou et al., 2006; Mercuri & Angelique, 2004; Wolmer, Laor, Dedeoglu, & Yazgan, 2005).

Disasters are frequently associated with both short- and long-term physical and psychological problems (Neria et al., 2008; Pynoos et al., 1998). Some of the immediate problems children face includes having to live in difficult and stressful environments and facing the facts of the reality of the catastrophe. Researchers have shown that PTSD and depression are sustained for many years after disasters or traumatic events (Basoglu et al., 2002; Chemtob, Nakashima, & Carlson, 2002; Najarian et al., 1996; Shaw et al., 1996; Shen & Sink, 2002; Winje & Ulvik, 1998; Wolmer et al., 2005).

Previous postdisaster research has demonstrated that the effects of traumatizing events endure for a long time (Chemtob et al., 2002; Winje & Ulvik, 1998; Wolmer et al., 2005). A study of childhood PTSD 6 months after Hurricane Floyd revealed that children clearly demonstrated the devastating effects of a natural disaster (Russoniello, Skalko, McGhee, Bingham-Alexander, & Beatley, 2002). Goenjian et al. (2008) found that even 6½ years after the earthquake in Armenia, adolescent children continued to experience depression. Among these adolescents, children who lost both parents were at a higher risk for depression. In a field study, Chemtob et al. (2002) explored the effects of living through a hurricane 1 year after the disaster. They found that children, including those who were previously treated, were still suffering from psychological problems, including generalized fear and problems associated with the death of a family member and physical injury (Chemtob et al., 2002).

Traumatized Children and Projective Techniques

Vizek-Vidovic et al. (2000) studied the effects of war trauma on children and argued that using children rather than adults, such as parents, as sources of information about the effects of trauma results in different findings. They suggest that parents tend to underestimate the effects of trauma on their children. Thus, it is important to focus on the reports of children when the effects of trauma in young children are of concern (Vizek-Vidovic et al., 2000). Yule (2001) and Falasca and Caulfield (1999) argued that traumatized children or children who have had various emotional problems have difficulty expressing their feelings and emotions. Woolgar (1999) stated that projective tests have been used reliably with children when the goal is to have children express their hidden feelings and emotions. When questions about an experienced trauma are asked explicitly, there is a possibility that children will not respond openly (Woolgar, 1999). Thus, when there is an absence of standardized tests and when it is emotionally difficult to handle being directly asked about the events experienced, controlled and consistent research can be planned by using projective techniques.

Some of the ways in which projective techniques are carried out in research is through the use of story completion, play, and children's drawings. Through the use of play and drawings, children can reflect their feelings of pain, fear, and helplessness (Gray Deering, 2000; Veltman & Browne, 2002). Not only do they express their feelings, they can also become active participants in the healing process (Falasca & Caulfield, 1999; Orr, 2007) Using projective techniques such as children's drawings has been reported to be safe and effective in helping children relieve pain (Malchiodi, 2001; Orr, 2007). Mercuri and Angelique (2004) noted that it is important to ask children about their postdisaster reactions and they need to be encouraged to express their feelings and thoughts verbally to ease their pain. This often helps children become more accepting of reality and facilitates a return to their normal lives (Shen & Sink, 2002). Previous research focusing on children's drawings after the Marmara disaster showed that children's drawings (with dark colors like black and brown) depicted disaster scenes including graves of parents or siblings (Celebi Oncu & Karaaslan, 2002). Another study carried out in the same disaster region with preschoolers' drawings showed that traumatized children tended to reflect their negative emotions about disastrous event in their drawings (Celebi Oncu, Akman, Guler, & Karaaslan, 2009).

Current Study

The goal of this exploratory study was to identify the consequences of the Kocaeli earthquake among children who were in kindergarten at the time the disaster struck. It was hypothesized that children who experienced the trauma of earthquake would still be feeling more negative thoughts 20 months after the disaster than children who did not experience trauma, when they are presented with a stressful situation. Because this study is one of the first of its kind using story completion as a projective technique to explore the inner workings of children experiencing disaster, the current study was designed as an exploratory study with a smaller sample size. The goal was both to understand the differences between the children's responses who experienced disaster and who did not and to explore whether a story completion as a projective technique could distinguish children's responses with exposure and no exposure to disaster.

To assess the types of negative and positive reactions young children would give to stressful situations, three different types of stories and three different types of questions were asked. In these stories, young children are presented with a scenario in which the main actor is facing some stressful event that shows similarities with the Kocaeli earthquake such as darkness, loud noise, shaking, trembling, and everything happening unexpectedly. The children were presented with incomplete stories and asked (a) how the story is going to end, (b) what the main character should do, and (c) how the child would end the stories. It was hypothesized that children who experienced trauma would give more negative responses to questions about how the story is going to end, and the children who did not experience trauma would give more positive responses and fight to change the situation. The second hypothesis was that the responses of the children who experienced trauma would report that the main character would be more likely to face trauma and injury, whereas children who did not experience trauma would be more likely to report survival of the main character and happiness. It was also hypothesized that if children are given a chance to end the stressful event themselves they would be less likely to end the stories negatively.

The specific predictions of the study were that the children who had experienced the earthquake disaster would complete the stories with negative answers more frequently than would children who had not experienced a disaster. Girls were expected to give more negative answers and complete the stories more negatively than boys as they were expected to be affected more by the disaster.

Method

Participants

The data for the current study were collected in May 2001. Two groups of children were included in this study. The first group of children (traumatized group) experienced the Kocaeli earthquake, while the second group (nontraumatized) did not. Because of the exploratory nature of the study a total of 103 children between the ages of 6.7 and 8.1 years (mean age = 7.4) at the time of the earthquake were selected to participate in this study. Participants in the traumatized group were selected from two elementary schools in the city of Golcuk, which is near the epicenter of the earthquake. The schools were selected because they were in the city center and most of the children in the schools were affected by the earthquake to some extent. Participants for the study group needed to meet two of three selection criteria. In addition to being physically present in the city of Golcuk when the earthquake hit, these criteria were to either lose a family member, to be physically injured themselves, or to experience destruction or serious damage to their homes. Out of these children, 84% of the children's houses were destroyed or seriously damaged, 23% had lost a family member, and 30% had survived the disaster with some physical injuries. The information about the types of trauma experienced by children was obtained from the families. At the time of data collection, the age of children in the traumatized group ranged from 8.4 to 9.9 years, with a mean of 9.1, and included 26 females and 27 males.

Children who were not victims of the earthquake were selected from the city of Kazan in Ankara Province. Kazan, which was 450 km away from the city of Golcuk, was selected, because the citizens living in this city did not experience the earthquake, and the two cities show some similarities in demographics. Even though Kazan, the city where the control group was selected, has a smaller population than the city of Golcuk, both cities have military bases; they both have industrial estates and are comparable in terms of their level of industrialization. The schools that were selected for the study are located in the areas where people come from middle socioeconomic status (Kazan and Golcuk municipalities). The elementary school in the city of Kazan was also located in the city center and there were a total of 450 primary school students enrolled in the school. A total of 50 children were selected to be in the control group. None of these children had experienced a life-threatening trauma by the time the data were collected. There were 25 females and 25 males in this group with ages ranging from 9.7 to 10.4 years, with a mean age of 9.8 years.

Materials

Three short and incomplete stories, which included words and descriptions used to remind children the earthquake, were prepared for this study (see Appendix A). The descriptions that were used to remind children of the earthquake were selected based on the general characteristics of the earthquakes, some of which are common for other natural disasters, and the specific features of Kocaeli earthquake. For example, the terms that were used in these stories to describe the event happening were "swinging and rocking back and forth, shaking, loud noises, things happening unexpectedly." In addition, because Kocaeli earthquake happened at 3:02 a.m., the stories included depictions of darkness. Even though stories used in the study included animals, plants, and inanimate objects, these aforementioned characteristics of an earthquake were used to describe the events which happened to the main characters in these stories. There were three short stories used in the study, one with a cat, the one with an ant, and a third with a tree as a main character. Each story was written in language appropriate for 9-year-old children. The teachers of children used in the study were given the stories first to determine whether the children could read and comprehend the stories without any help from their teachers. The children were read these stories and the questions associated with each story in their classrooms by one of the researchers (See Appendix B). These questions asked what happened to the main character, how the story ended and what the main character could do while the event was happening.

Procedure

The study was carried out in the children's elementary schools 20 months after the disaster. Data from the traumatized group were collected in April in two primary schools. Basic demographic information about the children was obtained from the children's parents. Although there were initially 30 boys and 30 girls in the study, 3 of the children refused to participate in the study and 4 did not meet the selection criteria. Thus, the responses of 53 children remained for analysis. The nontraumatized group was chosen as a comparison group from a primary school from a district of Ankara. The study was carried out in April with the traumatized group and in May with the nontraumatized group.

The children were assembled into groups of 10 to hear the stories. They were taken to a room where each child was seated separately, and each child was given a pencil and a sheet of paper. One of the authors remained with the children during the procedure. The researcher told the children:

> Now I'm going to read you some incomplete stories one at a time. Please listen carefully and when I finish each of these stories, I am going to ask you some questions about them. Please write your answers on the paper that is in front of you. You can take as long as you want. When you finish the first story, I am going to read the next story and we will repeat the same thing and answer the questions about that story.

Data Coding and Interrater Reliability

The authors, one of whom was completely blind to the children's situations including age, gender, and traumatic experience, coded the children's answers using the following procedure. First, both coders individually read children's answers and identified coding categories separately. Next, the coders met and shared the coding categories that they created based on the answers of the children. Similar codes were combined, and the codes that were different were examined and recreated by both coders. In the end, a total of seven coding schemas remained and the coders coded children's answers based on these coding schemas individually. After the first round of coding was completed, the agreement between the coders was established to be 88%, indicating that out of possible 824 answers, 97 of them did not match. Once the coders reviewed their answers and created one more coding category, the final category of codes reached to a total of eight. Once again the coders coded children's answers individually and in the end, when the answers of the coders were compared, it was found that the agreement between the coders reached to 96%, indicating a very good interrater reliability. To deal with the remaining 4% disagreement between the coders, those 34 responses of were examined by both coders cooperatively and an agreement was reached for 10 more items; the remaining 24 answers of the children that were either not clear, did not fit into any of the categories because of its content, or how it was written were labeled as missing. The coding schemas and their descriptions can be found in Table 1.

Statistical Analyses

For the statistical analysis, first the percentages of children whose answers fell for different categories were calculated for both traumatized and nontraumatized children and boys and girls for each of the three stories. Chi-square statistics

Table 1
Coding Schema and Children's Responses

Coding schema	General themes	Children's answers
Very negative	Death, destruction, serious injury, extreme distress	Killed; home was destroyed; train ran over the kitty; the roots of the tree came off and died; machine destroyed their home/neighborhoods; killed them . . .
Negative reactions	Distressed, scared, shaken, uncoordinated, upset, angry	Kitty got really scared; ant was afraid of the loud machine; kitty could not read his or her book; everything got dark; the leaves of the tree fell off; the ant was hungry . . .
Unsure	Not sure, trying to figure out what is going on, unaware	Kitty did not understand what was going on; looked outside to see what was going on; asked people to see what the noise was . . .
Sameness	Life goes on as it was before, unaffected, no traumatic consequence	Kitty went back to reading his or her book, nothing happened to the tree, did not fall, the machine just went away . . .
Good boy, good girl	Being calm, quiet, listening to what is said, following orders, being a good boy and a girl	Quietly waited for the noise/train/the machine to go away, wind to stop; listened to what is being said and obeyed; was a good kitty, did everything he/she was supposed to
Happiness	Happy ending, escaping from harm and danger, rescue, being safe	Kitty escaped, the tree/kitty/ant lived happily/very happily; ran away from danger for rescue . . .
Strength	Fighting against the destroyer, being strong, victory, building a new home	Stopped the train so that it would not destroy home/would not kill; stop the machine and kick it out of the forest; so strong stopping the wind . . .
Helping others	Caring for the others around, helping them, getting help/food for them	Went to look for siblings; rescued siblings and friends, brought food for others . . .

were reported to examine whether the numbers of children whose answers fell within different categories were statistically significant.

The next step in the analysis was to examine the differences between traumatized and nontraumatized children in terms of their negativity and positivity. To further the analyses, the answers of the children were recoded into a 4-point Likert scale for lower numbers to reflect high levels of negativity and high numbers reflect high levels of positivity. Even though questions asked for the stories were similar, because the characters and the events in the stories were different, the examination of the children for each of the story was carried out separately to investigate how each story made an effect on children's answers. To test this, three sets of multivariate analyses of variance (MANOVAs) were conducted.

Results

The first step in the analysis was to compare the number of children whose answers fell in different coding schemas based on child trauma status and gender. To examine the responses of children when they are presented with an incomplete story of a tree who is being shaken and tossed from side to side by a strong wind, cross-tabulations were computed, and the differences among children's responses were examined for statistical significance using chi-square analysis. When the children were asked about what happened to the tree in the story, it was found that traumatized children differed in their responses from children who were not traumatized, $\chi^2(3) = 8.72$, $p < .05$. It was observed that traumatized children were more likely to give either very negative responses (60%) such as the tree collapsing and its roots coming off or the tree dying, or negative responses (25%) such as the leaves falling, the three shaking form side to side, and feeling almost like it is going to come off from the ground. Even though the nontraumatized group was also likely to negative responses, fewer nontraumatized children gave very negative responses (38%) and more gave moderately negative responses (50%). Refer to Table 2 for a distribution of children's responses. The next common response for both the traumatized (15%) and nontraumatized (10%) children was that everything would stay the same, nothing would change, "the wind would probably shake the tree some, but, nothing would happen."

When children were asked about how they would end the story of the tree (see Table 3), differences between traumatized and nontraumatized

Table 2

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to the Question About What Happens to the Tree

	Trauma	No trauma	Total N
Very negative			
n	32	19	51
% within S1 Q1	62.7	37.3	100.0
% within trauma groups	60.4	38.0	49.5
Negative			
n	13	25	38
% within S1 Q1	34.2	65.8	100.0
% within trauma groups	24.5	50.0	36.9
Sameness			
n	8	5	13
% within S1 Q1	61.5	38.5	100.0
% within trauma groups	15.1	10.0	12.6
Happiness			
n	0	1	1
% within S1 Q1	0	100.0	100.0
% within trauma groups	0	2.0	1.0
Total			
n	53	50	103
% within S1 Q1	51.5	48.5	100.0
% within trauma groups	100.0	100.0	100.0

Note. No effects of gender were observed in the children's responses to what happens to the tree in the end, $\chi^2(3) = 6.06$, $p > .05$.

Table 3

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to the Question About How They Would End the Story About the Tree

	Trauma	No trauma	Total N
Very negative			
n	12	12	24
% within S1 Q2	50.0	50.0	100.0
% within trauma groups	22.6	24.0	23.3
Negative			
n	4	7	11
% within S1 Q2	36.4	63.6	100.0
% within trauma groups	7.5	14.0	10.7
Unsure			
n	1	0	1
% within S1 Q2	100.0	0	100.0
% within trauma groups	1.9	0	1.0
Sameness			
n	18	14	32
% within S1 Q2	56.3	43.8	100.0
% within trauma groups	34.0	28.0	31.1%
Happiness			
n	15	15	30
% within S1 Q2	50.0	50.0	100.0
% within trauma groups	28.3	30.0	29.1
Strength			
n	3	2	5
% within S1 Q2	60.0	40.0	100.0
% within trauma groups	5.7	4.0	4.9

children were not observed, $\chi^2(5) = 2.43$, $p > .05$, indicating that there were no differences between traumatized and nontraumatized children in terms of how positively or negatively they would end the story of the tree. Results suggested that numbers of children who would have negative or severely negative endings for the story of the tree were similar among traumatized (30%) and nontraumatized (38%) children. Moreover, among traumatized children 28% reported that they would have happy endings for the story, while among nontraumatized children it was 30% of the children. Again, children ending the story by indicating that everything would stay the same and nothing would change if they were given a chance to finish the story was again similar for traumatized (34%) and nontraumatized (28%) children. There was no difference between boys and girls, $\chi^2(5) = 7.10$, $p > .05$.

Next, children's responses to an incomplete story about a kitten who is threatened by a loud train coming closer was examined. In the first question, children were asked about what happens after the kitten looks from the window and sees the train and the electricity in the whole city is cut. Results revealed that children from traumatized group differed in their responses from the nontraumatized group, $\chi^2(6) = 15.46$, $p < .05$. It was found that more children in the traumatized group (86%) gave responses that were very negative than the no-trauma group (14%). When children's responses within trauma and no-trauma groups were examined separately, it was found that the majority of the traumatized children gave either negative (45%), or very negative responses (23%), whereas the majority of the nontraumatized children gave responses that were either moderately negative (39%) or positive (38%) with reports of strength, escape, and happiness. Very negative responses of the children included "the kitten would be hit by the train and die" and "the train would run over the kitten," while the negative responses included statements such as "the kitten would be scared from the loud noise, could not read his or her book, the electricity would stay cut." A number of nontraumatized children reported that everything would be the same (14%), "the kitten would continue reading, would look from the window and go back to what he or she was doing" (see Table 4).

Table 4

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to What Happens in the Kitten Story Next

	Trauma	No trauma	Total *N*
Very negative			
n	12	2	14
% within S2 Q1	85.7	14.3	100.0
% within trauma groups	22.6	4.1	13.7
Negative			
n	24	19	43
% within S2 Q1	55.8	44.2	100.0
% within trauma groups	45.3	38.8	42.2
Unsure			
n	3	8	11
% within S2 Q1	27.3	72.7	100.0
% within trauma groups	5.7	16.3	10.8
Sameness			
n	8	7	15
% within S2 Q1	53.3	46.7	100.0
% within trauma groups	15.1	14.3	14.7
Happiness			
n	2	10	12
% within S2 Q1	16.7	83.3	100.0
% within trauma groups	3.8	20.4	11.8
Strength			
n	1	1	2
% within S2 Q1	50.0	50.0	100.0
% within trauma groups	1.9	2.0	2.0
Good boy, good girl			
n	3	2	5
% within S2 Q1	60.0	40.0	100.0
% within trauma groups	5.7	4.1	4.9
Total			
n	53	49	102
% within S2 Q1	52.0	48.0	100.0
% within trauma groups	100.0	100.0	100.0

Table 5

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to How the Kitten Story Should End

	Trauma	No trauma	Total *N*
Very negative			
n	11	5	16
% within S2 Q2	68.8	31.2	100.0
% within trauma groups	21.6	10.2	16.0
Negative			
n	6	3	9
% within S2 Q2	66.7	33.3	100.0
% within trauma groups	11.8	6.1	9.0
Unsure			
n	0	2	2
% within S2 Q2	0	100.0	100.0
% within trauma groups	0	4.1	2.0
Sameness			
n	13	20	33
% within S2 Q2	39.4	60.6	100.0
% within trauma groups	25.5	40.8	33.0
Happiness			
n	18	16	34
% within S2 Q2	52.9	47.1	100.0
% within trauma groups	35.3	32.7	34.0
Strength			
n	1	1	2
% within S2 Q2	50.0	50.0	100.0
% within trauma groups	2.0	2.0	2.0
Good boy, good girl			
n	2	2	4
% within S2 Q2	50.0	50.0	100.0
% within trauma groups	3.9	4.1	4.0
Total			
n	51	49	100
% within S2 Q2	51.0	49.0	100.0
% within trauma groups	100.0	100.0	100.0

No gender effects were observed, $\chi^2(6) = 10.43$, $p > .05$.

Interestingly, however, when the children were asked about how the kitten story should end, no statistically significant differences between children from trauma and no-trauma groups were observed, $\chi^2(6) = 6.82$, $p > .05$ (see Table 5). The children who report a happy or positive ending for the story seemed to be similar, with 78% the nontraumatized children and 67% of the traumatized children (see Table 3). The happy ending responses included, "I would want the kitten to be happy, go back happily and read his or her favorite book, the kitten would be very happy to see the train." The responses that had strength and courage included "the kitten would stop the train so that the train would not hurt anybody." No sex differences were observed, $\chi^2(6) = 5.78$, $p > .05$. Refer to Table 2 for a distribution of children's responses.

The last question about the kitten asked children how they think the kitten should behave or respond. Even though there were some differences observed, they did not reach the levels of statistical significance set for the present study, $\chi^2(7) = 13.18$, $p = .07$. It is worth noting that only one traumatized child and no nontraumatized child gave a very negative response to this question, and only 2 traumatized children and 6 nontraumatized children gave somewhat negative responses. The majority of the children reported that the kitten should behave well (good boy/good girl) with 54% of the nontraumatized, and 43% of the traumatized children giving such responses (see Table 6).

The final story presented to the children in the study was about a hard-working ant who, while

Table 6

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to How the Kitten Should Behave/Respond

	Trauma	No trauma	Total N
Very negative			
n	1	0	1
% within S2 Q3	100.0	0	100.0
% within trauma groups	2.0	0	1.0
Negative			
n	2	6	8
% within S2 Q3	25.0	75.0	100.0
% within trauma groups	3.9	12.0	7.9
Unsure			
n	2	1	3
% within S2 Q3	66.7	33.3	100.0
% within trauma groups	3.9	2.0	3.0
Sameness			
n	3	8	11
% within S2 Q3	27.3	72.7	100.0
% within trauma groups	5.9	16.0	10.9
Happiness			
n	13	3	16
% within S2 Q3	81.2	18.8	100.0
% within trauma groups	25.5	6.0	15.8
Strength			
n	7	4	11
% within S2 Q3	63.6	36.4	100.0
% within trauma groups	13.7	8.0	10.9
Helping others			
n	1	1	2
% within S2 Q3	50.0	50.0	100.0
% within trauma groups	2.0	2.0	2.0
Good boy, good girl			
n	22	27	49
% within S2 Q3	44.9	55.1	100.0
% within trauma groups	43.1	54.0	48.5
Total			
n	51	50	101
% within S2 Q3	50.5	49.5	100.0
% within trauma groups	100.0	100.0	100.0

Table 7

Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to What Happens Next in the Ant Story

	Trauma	No trauma	Total N
Very negative			
n	18	6	24
% within S3 Q1	75.0	25.0	100.0
% within trauma groups	34.0	12.2	23.5
Negative			
n	12	7	19
% within S3 Q1	63.2	36.8	100.0
% within trauma groups	22.6	14.3	18.6
Unsure			
n	3	6	9
% within S3 Q1	33.3	66.7	100.0
% within trauma groups	5.7	12.2	8.8
Sameness			
n	0	1	1
% within S3 Q1	0	100.0	100.0
% within trauma groups	0	2.0	1.0
Happiness			
n	2	1	3
% within S3 Q1	66.7	33.3	100.0
% within trauma groups	3.8	2.0	2.9
Strength			
n	6	9	15
% within S3 Q1	40.0	60.0	100.0
% within trauma groups	11.3	18.4	14.7
Helping others			
n	11	19	30
% within S3 Q1	36.7	63.3	100.0
% within trauma groups	20.8	38.8	29.4
Good boy, good girl			
n	1	0	1
% within S3 Q1	100.0	0	100.0
% within trauma groups	1.9	0	1.0
Total			
n	53	49	102
% within S3 Q1	52.0	48.0	100.0
% within trauma groups	100.0	100.0	100.0

resting at home in the forest, hears a loud machine going through the forest and knocking down the trees. The first question asks children about what happens next. It was found that even though children reported various responses the difference between the responses of children from trauma and no-trauma groups was not at the levels of conventional statistical significance, $\chi^2(7) = 13.25$, $p = .07$. Of the traumatized children, 34% reported that something very negative would happen. These responses included the "ant dying, everything being torn down and nothing surviving and other people dying." Nearly 23% of them reported that something moderately negative would happen such as the ant would be scared, they could not find food, they would be upset, ant could not find his or her siblings. Nearly, 20% of the nontraumatized children reported helping others in need while only 11% of the traumatized children reported such behaviors. Over 18% of the nontraumatized children reported strength and courage, and over 11% of the traumatized children gave similar responses. These strength responses included "fighting against the machine," challenging it by yelling at it "to stop hurting the forest and the animals," "stopping it, kicking it out of the forest, building their homes and/or the forest back finding a lot of food to eat" (see Table 7). No gender differences were observed

Table 8
Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to How They Would End the Story About the Ant

	Trauma	No trauma	Total *N*
Very negative			
n	5	5	10
% within S3 Q2	50.0	50.0	100.0
% within trauma groups	9.8	10.0	9.9
Negative			
n	7	6	13
% within S3 Q2	53.8	46.2	100.0
% within trauma groups	13.7	12.0	12.9
Unsure			
n	0	1	1
% within S3 Q2	0	100.0	100.0
% within trauma groups	0	2.0	1.0
Sameness			
n	5	1	6
% within S3 Q2	83.3	16.7	100.0
% within trauma groups	9.8	2.0	5.9
Happiness			
n	17	17	34
% within S3 Q2	50.0	50.0	100.0
% within trauma groups	33.3	34.0	33.7
Strength			
n	13	12	25
% within S3 Q2	52.0	48.0	100.0
% within trauma groups	25.5	24.0	24.8
Helping others			
n	4	7	11
% within S3 Q2	36.4	63.6	100.0
% within trauma groups	7.8	14.0	10.9
Good boy, good girl			
n	0	1	1
% within S3 Q2	0	100.0	100.0
% within trauma groups	0	2.0	1.0
Total			
n	51	50	101
% within S3 Q2	50.5	49.5	100.0
% within trauma groups	100.0	100.0	100.0

Table 9
Cross-Tabulation of Traumatized and Nontraumatized Children With Different Responses to What the Ant Should Do

	Trauma	No trauma	Total *N*
Very negative			
n	2	0	2
% within S3 Q3	100.0	0	100.0
% within trauma groups	4.0	0	2.0
Unsure			
n	1	1	2
% within S3 Q3	50.0	50.0	100.0
% within trauma groups	2.0	2.0	2.0
Sameness			
n	0	4	4
% within S3 Q3	0	100.0	100.0
% within trauma groups	0	8.2	4.0
Happiness			
n	13	9	22
% within S3 Q3	59.1	40.9	100.0
% within trauma groups	26.0	18.4	22.2
Strength			
n	14	21	35
% within S3 Q3	40.0	60.0	100.0
% within trauma groups	28.0	42.9	35.4
Helping others			
n	15	9	24
% within S3 Q3	62.5	37.5	100.0
% within trauma groups	30.0	18.4	24.2
Good boy, good girl			
n	5	5	10
% within S3 Q3	50.0	50.0	100.0
% within trauma groups	10.0	10.2	10.1
Total			
n	50	49	99
% within S3 Q3	50.5	49.5	100.0
% within trauma groups	100.0	100.0	100.0

in children's responses, $\chi^2(7) = 8.72$, $p > .05$. Refer to Table 7 for a distribution of children's responses.

The next question asked children how they would end the ant story. No differences between the trauma and the no-trauma children were observed, $\chi^2(7) = 5.59$, $p > .05$. Similar to the story about the kitten, when children were asked how they would end the story, the negative responses were fewer than the positive responses (see Table 8). It was found that over 76% of the traumatized children reported that if it were them they would have a positive ending for the ant story, with nearly 10% reporting that everything would stay the same and life would go on as usual, over 34% reporting happy ending, nearly 26% reporting strength and courage to fight back, and nearly 8% reporting helping others. Interestingly, reports of positive endings for the story was also 74%, with 2% reporting that everything would stay the same, 34% reporting a happy ending, 24% reporting courage and strength, and 14% helping others. No significant gender differences were observed, $\chi^2(7) = 10.63$, $p > .05$.

When the children's responses to the final question about what the ant should do while the events were happening were examined, it was found that again, there were no significant differences between the trauma and the no-trauma groups, $\chi^2(6) = 9.62$, $p > .05$. Similar to the question about how they would end the story, the responses of children from both groups were positive. Of the traumatized children, 94% gave positive responses, with 26% happiness, 28% strength, 30% helping others, and

10% being a good boy or a girl. Almost all, over 98%, of the nontraumatized children gave positive responses, with over 8% reported doing the usual things, over 18% reporting happiness, nearly 43% reporting strength and courage, over 18% helping others, and over 10% reporting being a good girl and a boy. Refer to Table 9 for a distribution of children's responses.

The next step in the analysis included comparing traumatized and nontraumatized children in terms of their negativity and examining the interaction effects of trauma and gender. Three sets of MANOVAs were conducted to test the difference between children for each of the three stories. When children's negativity was examined for the story about a tree, multivariate analysis did not reveal a significant interaction effect, $F(2, 98) = 0.19$, $p > .05$; or main effects of gender, $F(2, 98) = 2.06$, $p > .05$; or trauma, $F(2, 98) = 1.06$, $p > .05$; nor were there any significant univariate effects of gender or trauma.

The next step in the analysis included assessing how negative or positive the children's responses were in response to the story about the kitty. Examination of the multivariate analysis did not reveal a significant interaction effect of gender and trauma, $F(3, 91) = 1.3$, $p > .05$. Multivariate analysis did, however, reveal a significant main effect of trauma, $F(3, 91) = 6.94$, $p < .05$, $\eta_p^2 = .19$, but no significant effects of gender, $F(3, 91) = 2.47$, $p > .05$, were found. Examination of the results of the univariate analyses did not reveal a significant interaction effects of gender and trauma for how the story ends, $F(1, 96) = 1.95$, $p > .05$; how the story should end, $F(1, 96) = 1.21$, $p > .05$; and how the kitty should behave, $F(1, 96) = 0.45$, $p > .05$. There was however, a main effect of trauma on children's responses about how the story ends, $F(1, 96) = 12.8$, $p < .05$, $\eta_p^2 = .12$, indicating that traumatized children ($M = 2.12$, $SE = 0.12$) ended the stories less positively than children who were not traumatized ($M = 2.73$, $SE = 0.12$). There was also a main univariate effect of trauma on children's responses about how the kitty should behave, $F(1, 96) = 6.71$, $p < .05$, $\eta_p^2 = .07$, indicating that traumatized children ($M = 3.36$, $SE = 0.09$) reported more positive responses than the nontraumatized children ($M = 3.04$, $SE = 0.9$). Another significant univariate finding was the significant main effect of child sex on how the kitty should behave, $F(1, 96) = 4.15$, $p < .05$, $\eta_p^2 = .04$, indicating that girls ($M = 3.08$, $SE = 0.9$) were less positive than boys ($M = 3.32$, $SE = 0.9$).

The final step in the analysis was to assess the negativity of children's responses when they were presented with the story of an ant with a loud machine in the forest based on their sex and whether they were traumatized. Multivariate analyses did not reveal an interaction effect of gender and trauma, $F(3, 91) = 1.39$, $p > .05$, or a main effect of gender, $F(3, 91) = 1.53$, $p > .05$. Multivariate analyses however, revealed a significant main effect of trauma, $F(3, 91) = 3.12$, $p < .05$, $\eta_p^2 = .09$. Examination of the results of univariate analyses indicated that child sex and trauma did not have significant interaction effects on for how the story ends, $F(1, 93) = 2.04$, $p > .05$; how the story should end, $F(1, 93) = 0.21$, $p > .05$, and what the ant should do, $F(1, 96) = 1.79$, $p > .05$. Further examination of the univariate analysis indicated that child trauma had a main effect on children's responses to how the story ends, $F(3, 91) = 12.91$, $p < .05$, $\eta_p^2 = .09$. Expectedly, traumatized children ($M = 2.53$, $SE = 0.17$) were more negative (less positive) than children who did not experience trauma ($M = 3.26$, $SE = 0.17$). Trauma, however did not have any univariate effects on how the story should end, $F(1, 93) = 0.19$, $p > .05$, and what the ant should do, $F(1, 93) = 0.02$, $p > .05$. The univariate effects of child sex was examined and no statistically significant effects were found for how the story ends, $F(1, 93) = 0.38$, $p > .05$, and how the story should end, $F(1, 93) = 1.21$, $p > .05$. Even though not significant at the levels of conventional significance, a trend toward statistical significance was observed for univariate gender effects on what the ant should do, $F(1, 93) = 3.78$, $p = .06$, $\eta_p^2 = .04$, indicating that girls ($M = 3.63$, $SE = 0.08$) demonstrated less positivity than boys ($M = 3.89$, $SE = 0.08$).

Discussion

This study examined the long-term consequence of experiencing a natural disaster. Traumatized children's and nontraumatized children's views about disastrous events appeared significantly different. Traumatized school-age children who witnessed or experienced destruction, grief, pain, and life threat, almost 2 years prior to the research, reported more negative thoughts about events that have themes similar to the trauma they experienced. This may be the consequence of the trauma, as the children could still be suffering postdisaster reactions. In a study carried on the same disaster region with preschoolers, when children were requested to tell stories about some emotional drawings, it was found that children's stories were mainly and significantly related to the disaster (Celebi Oncu et al., 2009).

In the present study, it was found that children who were directly affected by the earthquake could still be suffering significant traumatic effects 20 months after the event as they were more likely to have negative thoughts and end stories more negatively. This finding is similar with other studies of traumatic events (Chemtob et al., 2002; Najarian et al., 1996; Shaw et al., 1996; Shen & Sink, 2002; Winje & Ulvik, 1998; Wolmer et al., 2005). These studies showed that 1 or 2 years after disasters, children were still suffering trauma symptoms and they noted that children reported a fear of death or being physically hurt after disasters.

An important finding in this study is that in order to interpret the responses of children to trauma, it is necessary to distinguish between the children's anticipations about how the events will end and how children would like the event to end. Traumatized children demonstrated more negativity about the consequences of events when they felt they could not change the course of events. It appears that when they are given a chance to end the story themselves, "how would you end the story," they are more likely to have less negativity and more positive descriptions. Interestingly, for example, in the story of the tree, both traumatized and nontraumatized children were thinking that the end of the story, when it is outside of their control, would be negative. However, traumatized children were more likely to give severely negative responses than nontraumatized children. When the children were asked how they would end the story, an opportunity to change the course of events, no significant differences appeared to emerge between the trauma and no-trauma groups. In fact, the number of children giving a happier ending to the story increased in both groups. Similarly, in the story of the kitten, the majority of the traumatized children reported either negative or very negative responses while the nontraumatized children completed the stories either with positive or less negative responses. Again, when the children were asked about how they would end the story, the differences between the trauma and no-trauma groups disappeared. Even though there were slightly more positive responses among the no-trauma group, children from both groups were much more positive in their responses. Similarly, in the story of an ant, when children were asked how they would end the story, the responses were more positive overall and no differences emerged between the groups.

Another interesting finding that emerges from this study is that both groups of children were more negative when they were asked about a tree. This was clear both when they were asked to anticipate an ending for the story and determine an ending for the story as well as when their responses were analyzed independent of the type of the question that is asked to have children come up with an ending. Children's responses were less negative and more positive when the question was asked about the kitten and the ant. It is possible to speculate that children see trees differently from animals. Future studies should explore whether these findings about children's different responses to expected and desired endings are reflective of more general differences in life outlook between the traumatized and nontraumatized children.

In the present study, male and female children were expected to complete the stories differently. We had hypothesized that girls would show more negative effects from the disaster and that they would complete the stories with more negative and emotional explanations, and boys were expected to complete with more aggressive explanations. The results indicated that when children were asked about how the kitty should behave, the girls were more negative than the boys and they also reported more negativity when they were asked about how the ant should behave. Results of the present study did not show substantial gender effects to draw any conclusions. However, even though the differences were very small, nontraumatized boys were more negative than the nontraumatized girls. Attachment research has shown that young boys show more negative behaviors because they are more vulnerable than females to deprivation of warm affectionate maternal supportiveness (Honig, 2002). The present study did not assess the effects of the types of trauma children experienced. We do not know anything about these children's conditions at home to make and substantial conclusions about the gender effect. It may be one reason for why the findings were not consistent with the literature. Dewaraza, Sato, and Ogawa (2006) reported that tsunami-affected boys were more anxious and had more negative psychological reaction compared to girls. However, many previous studies have shown that traumatized girls demonstrated a significant increase in the levels of anxiety, stress, or depression in comparison with male children (Giannopoulou et al., 2006; Goenjian et al., 1995; Shaw et al., 1996; Vizek-Vidovic et al., 2000; Yule, 2001). More research on gender differences in response to traumatic experiences needs to be done to clarify the variables that may result in conflicting findings.

Implications and Future Directions

Traumatizing events are difficult for most children. The findings of this research indicated that traumatized children were still suffering from the earthquake disaster nearly 2 years later. They expressed more negative thoughts about disaster events in response to the story stems they heard. Although the study design was simple and was not confirmed by additional instruments, the uncompleted stories seemed to be an effective tool to expose the traumatized children's feelings about traumatic events. This was especially evident in considering the answers given by the contrast group compared with the traumatized group. While the children in the contrast group gave more positive answers, the traumatized children's answers tended to be more negative. These differences provide evidence for possible long-term effects of catastrophic disaster on children's thoughts and feelings.

Findings also showed that traumatized children preferred to complete the stories with happy endings if they felt that they are in charge. It could be hypothesized that traumatized children want to get rid of their stressful inner feelings and want happiness in their life again. They have positive desires, yet when they are asked about how things will be around them, they still feel that things will be negative. Thus, it is very important to support these children's feelings that they are in charge. These are hopeful findings for future resiliency and healing from trauma in children's lives. Disaster experiences may linger in the thoughts and ideas of traumatized children. Their positive responses, however, should give hopeful determination to counselors and therapists working in schools and with families that children have possibilities for resiliency and recovery that may be much more difficult for adults to achieve. Moreover, future research on sufferers of trauma, particularly natural disasters, should investigate the dynamics of learned helplessness among children.

Future research needs to include other projective techniques or more variations on incomplete stories that probe for children's responses to disasters. Therapists reading some children's very negative story endings will want to focus their efforts on easing children's traumatic symptoms such as PTSD, grief, anxiety, or depression and provide more reassurances and supports for healing.

Stories or children's literature can also be effective tools for children to recover. Children should be given chance to open their minds, to tell their stories, and to talk about their feelings, thoughts, and concerns about their trauma experience. Bibliotherapeutic books or stories could be used for children to recover and express themselves. Traumatizing events such as disasters are unexpected events. Prevention programs might focus on building up techniques of positive coping skills for children and families and increasing the levels of self esteem and decreasing the levels of learned helplessness.

This study showed that simply asking the question differently revealed such striking differences in children's responses. It is therefore important that future research designs allow for children to give their answers under different conditions in which we can really understand what they want, what they think will happen, what they would desire to happen, and what they think others expect to happen.

Study Limitations

Although the findings of this exploratory study are important for many reasons, there are still significant limitations of the study requiring future research. In fact, this is only amplified because of some of the interesting finding the study revealed. The narrative stories used for the study and the coding categories established were developed for the first time for the current exploratory study. Thus, there was no established validity for the stories or for the codes. Future research is needed to be conducted with Turkish children for cultural validity and with children from different cultures for construct validity. Another important limitation of the study is that we did not control for the types of trauma children experienced and how these experiences affected the responses children provided. We also do not know anything about the recovery and the healing process of the children and their families. Experiences of these children after the earthquake would influence them to have more positive or negative views about life. Another limitation of the study is not including the overall quality of the home environment in the study. Next, because of the exploratory nature of the study, the sample size was fairly small, thus preventing us from any generalizations and conclusions. On the other hand, it provides some interesting findings and directions for future research. Consequently, it is essential that future research be conducted with more children and children from different socioeconomic backgrounds and varying proximity to the earthquake and varying degrees and types of trauma experienced at the time of the earthquake would allow for obtaining more insightful information. Moreover, it is also important to use similar projective techniques for children with different types of trauma to understand the inner workings of young children who face disaster and trauma. Longitudinal studies need to be carefully designed to track down how children deal with trauma and under what conditions healing occurs.

References

Azarian, A., & Skriptchenko-Gregorian, V. (1998). Children in natural disaster: An experience of the 1988 earthquake in Armenia. Retrieved September 12, 2002, from http://www.aaets.org/article38.html

Basoglu, M., Salcioglu, E., & Livanou, M. (2002). Traumatic stress responses in earthquake survivors in Turkey. *Journal of Traumatic Stress, 15*, 269–276.

Bulut, S., Bulut, S., & Tayli, A. (2005). The dose of exposure and prevalence rates of post traumatic stress disorder in a sample of Turkish children eleven months after the 1999 Marmara earthquake. *School Psychology International, 26*, 55–70.

Celebi Oncu, E., Akman, B., Guler, T., & Karaaslan, T. (2009). A report on traumatised and nontraumatised children's human figure drawings reflecting emotional effects of disastrous conditions. *Australasian Journal of Trauma Studies, 2009-1*. Available at http://www.massey.ac.nz/~trauma/issues/2009-1/oncu.htm

Celebi Oncu, E., & Karaaslan, T. (2002, June). *Examining the drawings of traumatized children who experienced 17 August and 12 November disasters*. Poster presented at Symposium of Early Childhood Education in the 21st Century, Hacettepe University, Ankara.

Chemtob, C. M., Nakashima, J., & Carlson, J. G. (2002). Brief treatment for elementary school children with disaster-related posttraumatic stress disorder: A field study. *Journal of Clinical Psychology, 58*, 99–112.

Dewaraza, R., Sato, H., & Ogawa, T. (2006). Anxiety in tsunami-affected children in Sri Lanka measured by Revised Children's Manifest Anxiety Scale and Synthetic House-Tree-Person Test. *International Congress Series, 1287*, 74–78.

Durkin, M. S., Khan, H., Davidson, L. L., Zaman, S. S., & Stein, Z. A. (1993). The effects of a natural disaster on child behaviour: Evidence for posttraumatic stress. *American Journal of Public Health, 83*, 1549–1553.

Falasca, T., & Caulfield, T. J. (1999). Childhood trauma. *Journal of Humanistic Counseling Education and Development, 37*, 212–223.

Giannopoulou, I., Strouthos, M., Smithe, P., Dikaiakou, A., Galanopoulou, V., & Yule, W. (2006). Post-traumatic stress reactions of children and adolescents exposed to the Athens 1999 earthquake. *European Psychiatry, 21*, 160–166.

Goenjian, A. K., Pynoos, R. S., Steinberg, A. M., Najarian, L. M., Asarnow, J. R., Karayan, I., et al. (1995). Psychiatric comorbidity in children after the 1988 earthquake in Armenia. *Journal of American Academy of Child and Adolescent Psychiatry, 34*, 1174–1185.

Goenjian, A. K., Walling, D., Steinberg, A. M., Roussos, A., Goenjian, H. A., & Pynoos, R. S. (2009). Depression and PTSD symptoms among bereaved adolescents 6(1/2) years after the 1988 Spitak earthquake. *Journal of Affect Disorders, 112*(1-3), 81–84.

Gokben Hizli, F., Taskintuna, N., Isikli, S., Kilic, C., & Zileli, L. (2009). Predictors of posttraumatic stress in children and adolescents. *Children and Youth Services Review, 31*, 349–354.

Gray Deering, C. (2000). A cognitive developmental approach to understanding how children cope with disasters. *Journal of Child and Adolescent Psychiatric Nursing, 13*, 7–15.

Gurwitch, R. H., Silovsky, J. F., Shultz, S., Kees, M., & Burlingame, S. (2002). Reactions and guidelines for children following trauma or disaster. *Communication Disorders Quarterly*, 23, 93–99.

Honig, A. S. (2002). *Secure relationship: Nurturing infant/toddler attachment in early care settings*. Washington, DC: National Association for the Education of Young Children.

Jones, R. T., Frary, R., Cunningham, P., Weddle, J. D., & Kaiser, L. (2001). The psychological effects of Hurricane Andrew on ethnic minority and Caucasian children and adolescents: A case study. *Cultural Diversity and Ethnic Minority Psychology, 7*, 103–108.

Kilic, C., Aydin, I., Taskintuna, N., Ozcurumez, G., Kurt, G., Eren, E., et al. (2006). Predictors of psychological distress in survivors of the 1999 earthquakes in Turkey: Effects of relocation after the disaster. *ACTA Psychiatrica Scandinavica, 114*, 194–202.

Malchiodi, C. A. (2001). *Trauma and loss: Research and interventions article using drawing as intervention with traumatized children*. Retrieved November 21, 2005, from http://www.tlcinstitute.org/drawingintervention.html

McAdam Crisp, J. L. (2006). Factors that can enhance and limit resilience for children of war. *Childhood, 13*, 459–477.

Mercuri, A., & Angelique, H. L. (2004). Children's responses to natural, technological and Na-Tech disasters. *Community Mental Health Journal, 40*, 167–175.

Monahon, C. (1997). *Children and trauma: A guide for parents and professionals*. San Francisco: Jossey-Bass.

Najarian, L. M., Goenjian, A. K., Pelcovitz, D., Mandel, F., & Najarian, B. (1996). Relocation after a disaster: Posttraumatic stress disorder in armenia after the earthquake. *Journal of American Academy of Child and Adolescent Psychiatry*, *35*, 374–384.

Neria, Y., Nandi, A., & Galea, S. (2008). Post-traumatic stress disorder following disasters: A systematic review. *Psychological Medicine*, *38*, 467–480.

Orr, P. P. (2007). Art therapy with children after a disaster: A content analysis. *The Arts in Psychotherapy*, *34*, 350–361.

Pynoos, R., Goenjian, A. K., & Steinberg, A. M. (1998). Children and disasters: A developmental approach to posttraumatic stress disorder in children and adolescents. *Psychiatry and Clinical Neurosciences*, *52*, 129.

Russoniello, C. V., Skalko, T. K., McGhee, S. A., Bingham-Alexander, D., & Beatley, J. (2002). Childhood posttraumatic stress disorder and efforts to cope after Hurricane Floyd. *Behavioral Medicine*, *28*, 61–72.

Shaw, J. A., Applagate, B., & Schorr, C. (1996). Twenty-one month follow-up study of school age children exposed to Hurricane Andrew. *Journal of American Academy of Child and Adolescent Psychiatry*, *35*, 359–364.

Shen, Y. J., & Sink, C. A. (2002). Helping elementary-age children cope with disasters. *Professional School Counseling*, *5*, 322–330.

United Nations Children's Fund. (2001). Retrieved June 29, 2002, from http://www.unicef.org/turkey

Veltman, M. W. M., & Browne, K. D. (2002). The assessment of drawings from children who have been maltreated: A systematic review. *Child Abuse Review*, *11*, 19–37.

Vizek-Vidovic, V., Kuterovac-Jagodic, G., & Arambasic, L. (2000). Posttraumatic symptomatology in children exposed to war. *Scandinavian Journal of Psychology*, *41*, 297–306.

Winje, D., & Ulvik, A. (1998). Long-term outcome of trauma in children: The psychological consequences of a bus accident. *Journal of Child Psychology and Psychiatry*, *39*, 635–642.

Wolmer, L., Laor, N., Dedeoglu, C., & Yazgan, Y. (2005). Teacher-mediated intervention after disaster: A controlled three-year follow-up of children's functioning. *Journal of Child Psychology and Psychiatry*, *46*, 1161–1168.

Woolgar, M. (1999). Projective doll play methodologies for preschool children. *Child Psychology and Psychiatry Review*, *4*, 126–134.

Yule, W. (2001). Post-traumatic stress disorder in children and adolescents. *International Review of Psychiatry*, *13*, 194–200.

Appendix A

Story 1: A Story About a Tree

There was once a tree living happily on a hill with its roots reaching deep in the ground. It lived many days and nights, and had many beautiful days. One day on the top of the hill where the tree lived a strong wind began blowing. The branches of the tree were moving from side to side. The wind was so strong it felt like the roots of the tree were going to come off.

Story 2: Kitty in the Darkness

A little kitty was sitting happily in its cozy home looking at the pictures of its favorite book. All of a sudden she started hearing the noise of a train coming closer and the sound of the train was getting louder and louder. The kitty left her book and went to the window to look outside. Right at that moment, the power of the whole city was cut.

Story 3: The Ant and the Buldozer

There was once a hard working ant who lived in a forest. This ant was always carrying food home and preparing for the winter. Finally the winter came and the ant went to its home underground. One day a bulldozer came making loud noises, bothering the animals and knocking down the trees on its way to dig a big hole in the ground. The ant hears the noise and goes outside to its siblings to see what is going on.

Appendix B

Story 1
  Question 1: What happens to the tree in the story?
  Question 2: How would you end the story?
Story 2
  Question 1: What happened next?
  Question 2: How should the story end?
  Question 3: How should the kitty behave/respond?
Story 3
  Question 1: What happened next?
  Question 2: How should the story end?
  Question 3: What should the ant do?

Child Development, July/August 2010, Volume 81, Number 4, Pages 1176–1191

Tsunami, War, and Cumulative Risk in the Lives of Sri Lankan Schoolchildren

Claudia Catani
Bielefeld University and vivo

Abigail H. Gewirtz and
Elizabeth Wieling
University of Minnesota

Elizabeth Schauer
vivo

Thomas Elbert
University of Konstanz and vivo

Frank Neuner
Bielefeld University and vivo

This study examines the impact of children's exposure to natural disaster against the backdrop of exposure to other traumatic events and psychosocial risks. One thousand three hundred ninety-eight Sri Lankan children aged 9–15 years were interviewed in 4 cross-sectional studies about exposure to traumatic life events related to the war, the tsunami experience, and family violence. Symptoms of posttraumatic stress disorder, somatic complaints, psychosocial functioning, and teacher reports of school grades served as outcome measures. A global outcome variable of "positive adaptation" was created from a combination of these measures. Data showed extensive exposure to adversity and traumatic events among children in Sri Lanka. Findings of regression analyses indicated that all 3 event types—tsunami and disaster, war, and family violence—significantly contributed to poorer child adaptation.

The vast majority of data on the psychological impact of children's exposure to disasters come from developed countries (Norris et al., 2002). Nonetheless, the enormity of the power of disasters on life in the developing world requires a consideration of their impact on children's development and adjustment: Deaths from natural disasters in the developing world account for 95% of deaths from disasters worldwide (World Bank, 2006). Moreover, particularly in politically unstable regions, exposure to disaster often co-occurs with exposure to other traumatic events, such as armed conflict and family violence (e.g., Catani, Jacob, Schauer, Mahendran, & Neuner, 2008), and other psychosocial risks, including extreme poverty. Although difficult to accomplish, studying children in such contexts offers a special opportunity to understand the contribution of exposure to single-disaster incidents compared with other traumatic stressors and cumulative risk on children's functioning and development.

Data from four cross-sectional studies examining the functioning of Sri Lankan (primarily Tamil) schoolchildren over the period 2002–2006 are reported in this study. The time periods of these studies (i.e., prior to and following the tsunami of 2004) provide for snapshots of traumatic risk and functioning among a child population exposed to a

---

The data of this study were collected in the context of a cooperative humanitarian project of the Zonal Department of Education Vallikamam, the Shantiham counseling center, the German Technical Cooperation (GTZ), and the nongovernmental organization vivo. We are grateful to all children, teachers, counselors, and staff members in Sri Lanka who took part in the studies. The research conducted in these studies was supported by the DFG (Deutsche Forschungsgemeinschaft), "Ein Herz für Kinder" foundation, and vivo. The authors would like to thank Mahendran Kohila and Nadja Jacob for their special contribution to training of local counselors and supervision of data collection. In addition, we would like to thank Daya Somasundaram and Sundaram Divakalala from GTZ for supporting the research in Sri Lanka with their continuous professional and logistical input. Authors declare they have no conflict of interest. The corresponding author had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Correspondence concerning this article should be addressed to Claudia Catani, Department of Psychology, Clinical Psychology, and Psychotherapy, Bielefeld University, Post Box 100131, 33501 Bielefeld, Germany. Electronic mail may be sent to claudia.catani@uni-bielefeld.de.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0011

variety of highly stressful events, including war, natural disaster, and family violence. Next, we provide a conceptual developmental framework and rationale for the current research, as well as an overview of the context of war and natural disaster in Sri Lanka.

A Risk and Protection Framework for Children's Functioning

Studies, primarily undertaken in the developed world, have revealed the importance of understanding children's functioning in the contexts of environmental risks, vulnerabilities, assets, and protective factors or processes (e.g., Garmezy, 1993; Masten, 2001; Rutter, 1987; see Masten & Gewirtz, 2006, for a definition of terms). Of particular relevance is the observation that among children living in "high-risk" situations—that is, exposed to events that may potentially impair their health or development—risk factors tend to co-occur and to "pile up" (Sameroff, Seifer, Zax, & Barocas, 1987). For example, children who witness interparental violence are at significant risk for maltreatment (Edleson, 1999); children exposed to war or natural disasters are particularly vulnerable to "secondary adversities" such as homelessness, parental loss, and community violence (Pynoos, Steinberg, & Piacentini, 1999).

Given these observations, there has been an increased focus in the child development field on examining the effects of *multiple* risks rather than the nature or power of specific individual risks in the context of children's adjustment (e.g., Kraemer et al., 1997). Risk factors are considered to act cumulatively in the sense that the presence of more risk factors in a child's life is related to an increased likelihood of negative adaptation. Thus, studies of risk incorporate several possible different types of risk factors as predictor variables that are hypothesized to elevate the likelihood of undesirable outcomes in children's adjustment. Rutter (1979) found that the number of risk factors, more than any specific risk factor, accounted for increased risk of psychopathology in children. Subsequently, studies have shown consistent relations between risk and functioning, with increased risk associated with poorer functioning in developmental domains (e.g., Gutman, Sameroff, & Eccles, 2002).

This literature provides a valuable, broad "brush stroke" in demonstrating the influence of the quantity of risk factors to which a child is exposed, regardless of the type of risks. Also consistent with a risk and protection framework in children's development is the examination of outcomes that cut across developmental domains rather than being limited to a single domain (e.g., psychopathology). Thus, consistent with the literature indicating correlates of healthy development at different developmental stages, studies of risk and resilience have incorporated outcome measures that include social, academic, and psychopathology indicators (Masten & Coatsworth, 1998; Werner & Smith, 2001).

In contrast to a broader risk framework, studies specifically focusing on trauma more typically have investigated the impact of exposure to a single traumatic stressor or particular class of stressors (e.g., maltreatment, war, etc.) on individual child outcomes (e.g., the presence or absence of psychopathology; La Greca, Silverman, Vernberg, & Prinstein, 1996). As a result, this literature has seen relatively less focus on the broader context of children's development (e.g., Pynoos et al., 1999). However, somewhat in parallel to the broader risk literature described before, studies consistently have documented a dose–response effect with regard to the number, types, and severity of trauma events experienced by children. That is, the closer a child's proximity to a traumatic event, or the more events experienced, the more likely a child is to experience posttraumatic stress disorder (PTSD; e.g., Hubbard, Realmuto, Northwood, & Masten, 1995; Neuner, Schauer, Catani, Ruf, & Elbert, 2006). We could find no studies of trauma, however, whose outcome variable was children's overall adjustment rather than specific psychopathology. We could also find no studies examining the effects of exposure to multiple different traumatic events (i.e., events of a different trauma class) compared to exposure to a single event type (e.g., war or domestic violence alone).

Investigating the risk context is consistent with efforts to incorporate the broader environment (i.e., family, community, and society) in understanding influences on children's development (e.g., Bronfenbrenner, 1979). This risk and resilience and positive adaptation perspective may be especially important for studies in the developing world, where children's lives are so frequently marked by traumatic events and severe adversity, and where little data have been collected to describe children's adaptation in these contexts. Bronfenbrenner's (1979) bioecological systems theory adds to this developmental and contextual perspective by framing risk and resilience within multiple and interlocking systems operating in a child's life. The bioecological approach emphasizes that a child's own biology is a primary environment fueling his or her development. The interaction between factors in the child's developing biosystem, his or her immediate family and community environment, and the societal context both fuels and guides children's developmental pathways. Changes or conflict in any one layer will have ripple effects throughout other layers (Berk, 2000). To study a child's development, then, we must look not only at the child and her immediate environment but also at the interaction of the larger co-occurring environments across micro-, meso-, exo-, and macro-level systems (Berk, 2000). Below, we describe the context of this study, and outline a rationale for gathering basic information about both traumatic stressors and children's adaptation in a developing world context.

The Context of Risk in the Developing World: War and Natural Disaster in Sri Lanka

When studying risk and resilience in children affected by trauma, insights from developing world countries where children are often affected by a variety of profound risks to life, health, and development are of high importance. In particular, descriptive data are needed that not only describe the contribution of traumatic risks to adjustment but also allow for conclusions about the relation between the amount of various types of trauma exposure and adaptation in children. In this regard, Sri Lanka, where a major natural disaster was preceded by a long-lasting civil war, offers an unfortunately valuable opportunity to assess the larger context of risk in childhood. Specifically, this context enables us to investigate the influence of exposure to different types of traumatic stressors (i.e., multiple risks), as well as the relation between the extent of exposure to each stressor type and adaptation.

In Sri Lanka's North-Eastern provinces, in particular, traumatic experiences were already common prior to the tsunami in December 2004. For two decades, thousands of people had been killed and many more were displaced as a result of conflict and violence in the country. The Liberation Tigers of Tamil Elam (LTTE) have been fighting for political autonomy resulting in ongoing conflict in the northern and eastern areas of the country. Civilians have become casualties of war, landmines and unexploded ordnance pose a constant threat, hospitals and schools have been destroyed, and the economy has been shattered. In spite of a cease-fire that was called in December 2001, violent incidents between the LTTE and the government have recently dramatically intensified, costing the lives of several thousand people over the last 2 years.

Since early 2002, our workgroup had been partnering with the German Technical Association, the Ministry of Education, the Education Council North-East, and the Shantiham Association for Health and Counseling to build up a cascade-model of psychosocial care for affected children in the North-Eastern provinces of Sri Lanka. The framework of this model consisted of over 1,000 teachers being trained to identify children with psychological impairments in schools. A selection of 150 of the most qualified teachers were trained as counselors to provide services to children in need. Over this period of time, four epidemiological surveys were carried out with Tamil school children, two studies prior to and two subsequent to the tsunami. The overall aim of these studies was to establish the severity of war-related (and later also tsunami-related) exposure to trauma in Tamil children and its impact on their mental health.

The first study was conducted in the Vanni area very shortly after the ceasefire agreement at the end of 2001 and showed that 92% of the children had experienced traumatizing war-related events and 25% of the sample met the criteria for a PTSD diagnosis (Elbert et al., 2008). These results were confirmed by a similar follow-up study in the Vallikamam and Jaffna educational zones 2 years later (Catani et al., 2005), where 80% of the interviewed school children had been traumatized by the war. The PTSD prevalence rate for children was 18% with the "traumatic load" (number of experienced event types) being significantly linked to the development of PTSD.

This finding led us to use the so-called dose–effect in further conceptualizing and understanding the impact of traumatic exposure on mental health (Neuner et al., 2004). This association was further established in the third epidemiological survey carried out 4 weeks after the tsunami hit Sri Lanka in December 2004. Results showed that the prevalence of PTSD in Sri Lankan children affected by both the civil war and a natural disaster was as high as 40% after the tsunami with PTSD rates being best explained by the severity of exposure to tsunami-related traumatic events and previous war experiences. To further extend the assumption of a dose–effect model to other types of stressors in the lives of children, in a subsequent study we included experiences of family violence as a potential risk factor for mental health impairments in Sri Lankan children (Catani, Jacob, et al., 2008). Apart from presenting a high prevalence of abusive experiences at home, results from this study showed that family violence, experienced or witnessed by children, ought to be taken into account when investigating the cumulative effects of various forms of stressful experiences on children's well-being.

The finding about the critical impact of family violence on children's mental health encouraged us to reanalyze all the data from the previous study and to take into consideration information about the experience of family violence, which had been assessed but was not included in previous data analysis. Furthermore, we wanted to not only focus on PTSD as the outcome measure but also to take into account the variety of outcome variables that had been assessed in the surveys (i.e., somatic complaints, school grades etc.) and combine them into an index for children's overall adjustment toward risk.

In summary, we broaden the investigation of the impact of traumatic events on children's functioning by applying a risk framework to the study of children's exposure to traumatic events in a highly stressful developing world arena. In this context, we were interested in (a) understanding the nature and influence of different types of traumatic stressors (risks) on children's functioning and (b) understanding children's positive adaptation in the broader developmental and environmental context (i.e., academic and social and health functioning, as well as absence of psychopathology).

Research questions related to the analysis of the four data sets were as follows:

1. How are different stressors associated with children's adaptation following exposure to war or tsunami?
2. Does assessment of a broader, more detailed range of risks/traumatic events result in more variance explained in children's adaptation?
3. Are both the multiplicity of risk (defined by the number of different risk categories a child is affected by) and trauma severity (defined by the number of single event type within a specific risk category) predictive of children's adaptation?

This last question, aimed at disentangling severity and the extent of different trauma types and multiplicity of risk, was based on DataSet 4 only, as this was the only study where continuous measures for risk and trauma severity were used.

Method

Samples and Procedures

Data presented here are taken from four independent epidemiological surveys that our workgroup conducted at different time points in different geographical areas of Sri Lanka. Details of the single surveys with respect to both methods (sampling strategy, data collection, instruments) and results have been published elsewhere (Catani, Jacob, et al., 2008; Catani et al., 2005; Elbert et al., 2008; Neuner et al., 2006), and will therefore not be addressed in full detail in the present study.

All studies were designed as surveys of school-attending children or—in the case of the Tamil children interviewed right after the tsunami—of school-aged children in a temporary shelter for internally displaced people. All participants were randomly selected from the respective populations. Within the schools included in Surveys 1, 2, and 4, participants were randomly selected on the basis of an index including classes with children of a certain age range. For every class, every other to every fifth child (depending on the size of the class) on an alphabetically ordered list was chosen and invited to participate in the survey. The study population in the immediate aftermath of the tsunami (Study 4) also included a sample of children recruited from a provisional refugee camp in Manadkaddu. All eligible children aged 8–14 who were present on the day of the interview were included in the survey. The four samples that were included in the present data analysis comprised a total of 1,398 children from different parts of Sri Lanka, mainly the Tamil North-Eastern region of the country.

Study 1 (Vanni). In Study 1 (Elbert et al., 2008), fifth-grade children from 15 different schools randomly chosen from all five administrative zones of the Vanni region of Sri Lanka participated in the survey. Vanni is a region under the control of the LTTE enclosed on the east and west by the sea and on the north and south by no-man zones at military lines that were established by the ceasefire agreement signed in 2002 between the government of Sri Lanka and the LTTE.

Sample 2 (Vallikamam). Study 2 (Catani et al., 2005) was conducted with children of six different schools of the Vallikamam and Jaffna educational zones in the north of Sri Lanka.

Sample 3 (Manadkaddu, Galle, and Kosgoda). Study 3 (Neuner et al., 2006) was carried out 1 month after the tsunami, in January 2005, in three severely affected communities of the country. The village Manadkadu in the Vadamarachi region in northern Sri Lanka was completely destroyed by the tsunami. Like many communities in the North-Eastern provinces, Manadkadu had been severely affected by the Sri Lankan civil war already, prior to the natural disaster. In contrast to the other surveys comprised in this study, Study 3 included not only children of the Tamil population of the North-East of Sri Lanka but also two locations in the Singhalese part of the country that had not been affected by the civil war. However, both regions—Kosgoda (western Sri Lanka) and Galle (southern Sri Lanka)—are coastal areas, and were both devastated by the tidal wave.

Sample 4 (Point Pedro). The most recent study (Catani, Jacob, et al., 2008) was conducted approximately 1 year after the tsunami in the Vadamarachi school district (Point Pedro area) at the northern tip of Sri Lanka. Being an important strategic point during wartime located right by the sea, Point Pedro had been severely exposed to two decades of civil war before it was hit by the natural disaster at the end of 2005. Table 1 provides an overview on location and time period of data collection in the four studies and on sociodemographic characteristics of the different child samples.

Procedures. In all surveys, informed consent was obtained from both the child and one of her caregivers or parents. Approval for implementation of the studies in Sri Lanka was obtained from the Ethical Review Board of the University of Konstanz, Germany. All surveys were carried out in cooperation with and approved by Sri Lanka's Ministry of Education. Interviews were conducted by former schoolteachers who had previously been trained as "master counselors" (Schauer et al., 2005) to assist traumatized children. Interviews conducted by local counselors were closely supervised by one or more clinical psychologists from our work group.

Measurement

Following the hypotheses derived from the theoretical background, we sought to measure a variety of risk factors including economic deprivation, loss of family members, family violence, violence associated with the war, traumatic events associated with the tsunami, and other traumatic experiences in each study. Psychosocial adaptation was the common outcome measure of the analyses. We created an adaptation index based on the following variables: somatic symptoms, PTSD symptoms, psychosocial functioning, and school performance. However, data presented here are a reanalysis of four independent studies where parts of the same and different instruments and outcome measures had been used. Table 2 provides an overview of independent variables and outcome measures for each survey to further clarify areas of methodological overlap versus differences between the studies.

Economic status. To reduce the amount of variables in the analyses, we constructed a single global measure of economic status from data related to economic factors in the surveys. The index of economic status was calculated as the mean of standardized (z-transformed) variables related to child possessions, family possessions (assessed using an asset checklist of typical items), and meals per week.

Family loss. In Studies 1, 3, and 4, the number of core family members who died was assessed by asking "How many of your family members, including your parents and siblings, have died so far?" In Study 2, only the loss of parents was assessed by a yes or no answer to the question "Did either your father or your mother die?"

War experiences. In Studies 1 and 4, war experiences were quantified using a war-related event checklist (Catani, Jacob, et al., 2008). In Studies 2 and 3, the event checklist of the University of California at Los Angeles (UCLA) PTSD Index (UPID) for children (Pynoos, Rodriguez, Steinberg, Stuber,

Table 1
Sociodemographic Data of the Four Independent Sri Lankan Samples

Location of survey	Sample 1 Vanni	Sample 2 Vallikamam	Sample 3 Manadkaddu, Galle, and Kosgoda	Sample 4 Point Pedro
Time of survey	2002	2004	2005	2005–2006
Months since tsunami	n.a.[a]	n.a.[a]	1	11–13
No. children interviewed	364	474	264	296
Gender (male) %	47.5	42.5	50.8	53.4
Mean age (range)	10.5 (10–14)	13 (10–16)	11.1 (8–15)	12.2 (9–15)
Loss of a parent (%)	n.a.[b]	13.5	13.3	10.1
Mean of lost core family members (range)	0.13 (0–2)	n.a.[b]	0.37 (0–4)	0.5 (0–9)
Mean number of possessions (*SD*)	27.1 (15.1)	n.a.[b]	n.a.[b]	57.4 (28.7)
Mean number of meals per week (*SD*)	18.4 (4.6)	n.a.[b]	n.a.[b]	20.1 (1.8)

[a]Not applicable. [b]Not assessed.

Table 2
Independent and Outcome Variables for Each DataSet—Overlap and Differences

	Sample 1	Sample 2	Sample 3	Sample 4
Independent variables				
Family violence	n.a.[a]	Sum of UPID E5 + E6	Sum of UPID E5 + E6	Number of different family violence types (checklist)
Family loss	Number of lost family members	Death of a parent (dichotomous item)	Number of lost family members	Number of lost family members
War exposure	Number of different war events (checklist)	UPID E4 (dichotomous item)	UPID E4 (dichotomous item)	Number of different war events (checklist)
Tsunami severity	n.a.[b]	n.a.[b]	Tsunami severity (checklist)	Tsunami severity (checklist)
Index for multiplicity of risk categories	n.a.[b]	n.a.[b]	n.a.[b]	Number of different risk categories (war, family violence, tsunami, loss of a parent) experienced
Outcome variables				
Physical health	Number of somatic complaints reported on the checklist for physical complaints	Number of somatic complaints reported on the checklist for physical complaints	Number of somatic complaints reported on the checklist for physical complaints	Number of somatic complaints reported on the checklist for physical complaints
School performance	Mean of school grades for math, science, English, and Tamil	Mean of school grades for math, science, English, and Tamil	Mean of school grades for math, science, English, and Tamil	n.a.[b]
Mental Health (PTSD severity)	UPID score	UPID score	UPID score	UPID score
Functioning (in three areas of life)	Number of "yes" answers to three questions about the presence of functioning impairments	Number of "yes" answers to three questions about the presence of functioning impairments	Number of "yes" answers to three questions about the presence of functioning impairments	Number of "yes" answers to three questions about the presence of functioning impairments
Adaptation Index	Mean of *z*-transformed values of physical health, school performance, mental health, and functioning	Mean of *z*-transformed values of physical health, school performance, mental health, and functioning	Mean of *z*-transformed values of physical health, school performance, mental health, and functioning	Mean of *z*-transformed values of physical health, mental health, and functioning

Note. PTSD = posttraumatic stress disorder; UPID = University of California at Los Angeles (UCLA) PTSD Index.
[a]Not assessed. [b]Not applicable.

& Frederick, 1998) was employed to assess exposure to the war. Previous exposure to the war was ascertained with a dichotomous measure to which children answered "yes" or "no" to the UPID event checklist item "being in a place where a war was going on around you" (Item E4).

Family violence. Based on the available data of each survey, in Studies 2 and 3, family violence was assessed by children's responses to the two UPID event checklist items referring to family violence: "being hit, punched, or kicked very hard at home" and "seeing a family member being hit, punched or kicked very hard at home" (Items E5 and E6). In the Point Pedro sample (Study 4), the amount of family violence was defined as the number of questions answered positively on the family violence event checklist. In the latter study, an additional 30-item questionnaire was utilized to assess the quantity of adverse childhood experiences at home (Catani, Jacob, et al., 2008). It is important to note that the checklist about family violence included a variety of experiences ranging from slightly aversive events occurring at home (e.g., being threatened verbally) to experiences that fall in the range of traumatic life events (e.g., having been threatened with a weapon).

Severity of exposure to the tsunami. To estimate the severity of tsunami exposure in the two post-tsunami samples, five questions related to the tsunami experience were added to the interviews (Catani, Jacob, et al., 2008; Neuner et al., 2006). For instance, children were asked "Did you see the big wave close by?" and "Were you caught by the wave?" The five questions were answered "yes" or "no." The score for objective Tsunami exposure was the number of "yes" answers.

Exposure to other traumatic experiences. To obtain a more complete representation of traumatic events as potential risk factors in the lives of Sri Lankan children in the analyses, we included additional items of the UPID event checklist where available. These items could not unambiguously be assigned to the context of war, family, or disaster, and could also be related to community violence. The items related to physical attack ("being beaten up, shot at or threatened to be hurt badly in your town"), witnessing physical attack ("seeing someone in your town being beaten up, shot at or killed"), having seen a dead body ("seeing a dead body in your town"), or experiencing sexual violence ("Having an adult or someone much older than you touch your private sexual body parts when you did not want them to").

Multiplicity of risks. In Study 4, in addition to the continuous measures of severity of a specific risk or trauma type, we created an index of multiplicity of risk defined by the number of different risk categories a child had been or was currently exposed to. Risk categories that could be determined were the experience of war, the experience of family violence, the experience of the tsunami, and the loss of a parent. On the basis of children's answers to the UPID event questions for war and for family violence, children were classified as either being affected by the specific risk category or not. With respect to the tsunami, children were classified as "having been exposed to the tsunami" if their tsunami severity score was > 0.

Adaptation score. A global adaptation score was calculated as the mean of z-transformed values of the following outcome measures that had been included in the four surveys (descriptive statistics with respect to the single outcomes, such as PTSD symptom score or number of somatic problems, will not be presented here as these results have been separately published elsewhere for each dataset; see indicated references). The means and standard deviations for the z-transformation of each measure were calculated using the scores from all four data sets. For Physical Health, Functioning, and Mental Health Indices, z values were transformed into negative values so that greater values would indicate better adaptation.

Physical health. Physical health was assessed using six questions about the presence of specific somatic complaints in the last 4 weeks. The complaints were headache, cough, pain, fever, vomiting, and diarrhea. For instance, the child was asked "Have you had diarrhea in the last 4 weeks?" and the item was coded "yes" or "no" according to the child's answer. The outcome measure for physical health was defined as the number of "yes" answers to the question about somatic complaints.

Mental health. The UPID for children (Pynoos et al., 1998) was used for the assessment of symptoms related to PTSD. A validation of the diagnostic instrument showed a high agreement in the PTSD diagnosis (79%) between the UPID as administered by the counselors and expert interviewers (Elbert et al., 2008). The UPID allows for the assessment of symptom severity based on the frequency of symptoms reported by the child. For each single *DSM–IV* symptom of PTSD, its occurrence within the last month is scored on a scale ranging from 0 (*none of the time*) to 4 (*most of the time*). Thus, an overall trauma severity score can be calculated by summing the scores for each question that corresponds to a symptom resulting in a maximum possible score of 68. Mental health was defined as the average trauma severity score of the UPID.

Functioning. Three simple questions related to problems in functioning in different areas of children's life (social relationships, family life, and general life satisfaction) were added to the interview. For instance, with respect to social relationships, the child was asked "In the past 12 months, have you lost your interest in playing or socializing with your friends?" If the child did not understand the meaning of the question, the interviewer was instructed to help the child with further questions or examples. Impairments in psychosocial functioning were quantified as the number of "yes" answers to the three questions.

School performance. Children's performance at school was defined as the average grade achieved in four classes: mathematics, science, Tamil, and English. School grades for each child were obtained from teacher's reports about the previous school term. We did not create an index of school performance for the children of Study 3, which was conducted only a month after the tsunami. As the available reports about school grades referred to the period prior to the tsunami, the inclusion of these outcomes would not have reflected the reaction to the disaster. Therefore, z-transformation of school grades was based on the means and standard deviations of the data sets from Studies 1, 2, and 4.

Translation of Instruments

Overall, Tamil language versions of questionnaires and standardized instruments were created using translation and blind back-translation; that is, a professional counselor translated from English to the respective local language, and independent translators performed the back-translation, blind to the original version. Discrepancies were checked by experts and final versions were derived through extensive consultation with local counselors and a team of translators.

Results

Descriptive Statistics

Exposure to the war. Exposure to war-related event types was very high in all samples of Tamil children. Eighty-nine percent of the children in Study 1 and 80% of children in Study 2 reported having experienced at least one war-related life event. With respect to the children interviewed right after the tsunami, there was a clear difference between Singhalese children living in the south of the country that had hardly been affected by the civil war, and the Tamil children in Manadkaddu. Only 4.7% of the 188 Singhalese children had been affected by the war, whereas 59.2% of the 72 Tamil children reported at least one war-related life event. In the 2006 sample, this proportion was even higher, with 82.4% of the children affected by war experiences. Children in Study 1 reported an average of 2.9 (SD = 2.2) different war-related event types, and children in Study 4 reported 2.0 (SD = 1.8) types of such experiences.

Experiences of family violence. In Studies 2–4, children were asked about having experienced or witnessed violence at home. In the Vallikamam sample, 24.1% of the children reported having either experienced violence by a family member or witnessed somebody being beaten at home. In Study 3, this percentage was similarly high, with 25.4% of the children. In the Point Pedro sample, children were presented with a detailed list of family violence events of varying severity. Here, almost all children (95.6%) reported at least one aversive event happening at home. The most frequent experiences were being slapped (81.8%), being hit with an object (68.9%), and seeing a family member being hit (55.4%). Another 4.3% of children reported having experienced or witnessed at least one incident of sexual abuse or sexual violence at home. On average, children in this sample reported to have experienced or witnessed 5.3 (SD = 3.2) different violent event types in their families.

Tsunami exposure. Of the children interviewed 4 weeks after the tsunami in Study 3, 87.9% had been directly affected by the disaster. Among these children, 87.1% had to flee from the wave, 34.9% had been caught by the wave, and 65.5% had seen people struggling or dying in the water. Among the 70.9% of the children in Study 4 who were directly affected by the tsunami, these percentages were 98.6%, 22.9%, and 40.5%, respectively. On average, the tsunami severity score was 2.9 (SD = 1.2) in Sample 3 and 2.4 (SD = 1.28) in Sample 4 on a scale ranging from 1 to 5.

Multiplicity of risks. Of the 296 children of the 2006 survey, only 4 (1.4%) had not been exposed to a single risk category. Whereas a small number of children (11.1%) were found to be affected by only one risk category, the majority of children (51%) were exposed to two risk categories. One third (33.1%) of the children were affected by three risk categories, and the remaining children (3.4%) in the sample were exposed to traumatic events in all four categories.

Adaptation score. Figure 1 provides an overview on the z-transformed outcome measures and the mean adaptation score calculated for each sample. To allow for better comparison between the independent child samples, we included only data from Tamil children (n = 71) from the North-Eastern provinces of the country and left out the Singhalese children (n = 193) interviewed in Study 3. It is important to note that we had to deal with a certain amount of missing data, in particular with respect to Surveys 1, 2, and 3, stemming from our inability to send interviewers back to schools to talk to children obtain missing information. For missing data, we adopted a rather strict method of excluding the missing item as well as the related score (e.g., PTSD symptom score) from further data analysis. As a result, we could not calculate the z-transformed outcomes and the global adaptation score for all children. The percentages of children included in the regression analysis for each data set are reported in Tables 3 and 4.

As shown in Figure 1, children of the Manadkaddu sample interviewed shortly after the tsunami

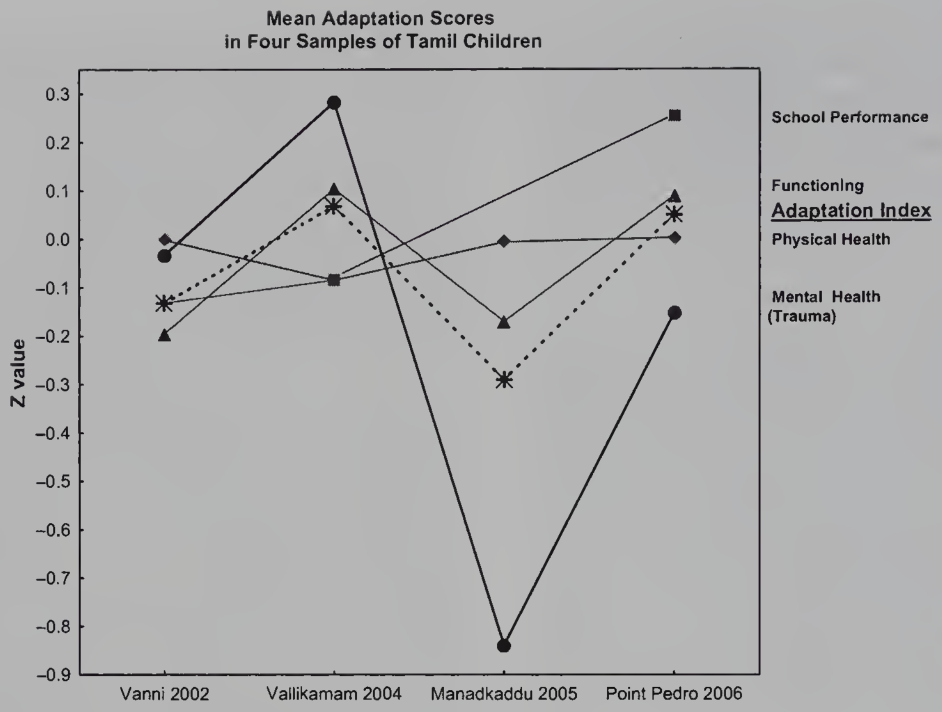


Figure 1. Plotted means of the z-transformed outcome measures (school performance, functioning, physical health, and mental health) and mean adaptation index for each dataset.
Note. ■ = school performance; ▲ = functioning; = adaptation index; ♦ = physical health; ● = mental health.

Table 3
Linear Regression Analyses on Adaptation Score (for Samples 1 and 2)

Predictor	β	*p* value	Zero-order correlation
Vanni 2002 (*N* = 270; 74% included), R^2 = .09; *F*(5, 175) = 4.57, *p* < .001			
Age	−.16	.03	−.18
Gender (male)	−.06	.37	.05
Economic status	.20	.007	.22
No. lost family members	−.08	.27	−.05
Previous exposure to war	−.17	.015	−.16
Vallikamam 2004 (*N* = 397; 84% included), R^2 = .27; *F*(10, 385) = 15.40, *p* < .0001			
Age	−.15	.001	−.19
Gender (male)	−.04	.31	.07
Loss of a parent	−.05	.23	−.08
Degree of family violence	−.26	< .0001	−.33
War experience	.06	.20	.06
Experience of a natural disaster	−.17	.0002	−.29
Physical attack (beaten or shot)	−.12	.008	−.24
Witnessing physical attack	−.21	< .0001	−.29
Having seen a dead body	−.08	.07	−.21
Experience of sexual violence	−.07	.11	−.13

showed the lowest scores with respect to mental health (trauma), functioning, and global adaptation.

Predicting Children's Adaptation

To examine potential predictors for adaptation, we calculated a linear regression model for each data set. The dependent variable is the mean adaptation score. Depending on what kind of independent variables had been assessed in the particular sample, different types and numbers of possible predictors were included for the different data sets. Results of the analyses for the two pretsunami samples are presented in Table 3, and results for the two independent samples that were assessed after the tsunami are presented in Table 4.

In the regression analysis for the 2002 study conducted in the LTTE area (Vanni), economic status and previous exposure to the war (inversely related) were the strongest predictors for children's adaptation. However, as few predictor variables were assessed in this first Sri Lankan survey, the full model only explained 9% of the variance in the adaptation index. With respect to the 2004 study, the inclusion of further traumatic events as potential predictors in the regression model resulted in 27% of variance explained. In this study, experiences of family violence, natural disaster, and physical attacks were the strongest negative predictors of adaptation. In both Studies 1 and 2, age had significant predictive power, with younger children showing higher adaptation scores.

In the regression analysis for the studies conducted after the tsunami, the severity of the natural disaster together with prior exposure to the war was among the strongest predictors of children's adaptation. In the study conducted in 2005, 1 month posttsunami, the number of lost family

Table 4

Linear Regression Analyses on "Adaptation Score" (for Posttsunami Samples 3 and 4)

Predictor	β	*p* value	Zero-order correlation
Posttsunami 2005 (*N* = 194; 74% included), R^2 = .30; *F*(10, 16) = 8.42, *p* < .0001			
Age	−.03	.65	−.22
Gender (male)	−.06	.32	−.04
No. lost family members	−.18	.008	−.33
Degree of family violence	−.10	.15	−.17
War experience	−.24	.0005	−.35
Physical attack (beaten/shot)	−.10	.13	−.16
Witnessing physical attack	−.06	.38	−.20
Having seen a dead body	−.14	.04	−.32
Experience of sexual violence	−.09	.19	−.17
Tsunami severity	−.21	.002	−.39
Point Pedro 2006 (*N* = 294; 99% included), R^2 = .26; *F*(7, 286) = 15.75, *p* < .0001			
Age	−.01	.86	−.04
Gender (male)	.05	.29	−.08
Economic status	.20	.0002	.32
No. lost family members	−.09	.07	−.09
Amount of family violence event types	−.23	< .0001	−.27
Amount of war event types	−.14	.01	−.25
Tsunami severity	−.32	< .0001	−.33

members was significantly related to a lower adaptation score in children. In the 2006 study (Study 4) where family violence and economic status were included as independent measures, the amount of violence experienced at home was inversely related to the adaptation score whereas economic status showed a strong positive relation with children's total adaptation. In both Studies 3 and 4, the regression model explained a significant amount of the variance, 30% and 26%, respectively, for each data set.

To further investigate the relation between the amount or severity of different risk factors (types of traumatic experiences related to war and family violence, tsunami severity, and loss of family members) and children's positive adaptation, we created subgroups of children from the two posttsunami surveys based on the number of stressful event types asked for in the UPID event checklist. Aversive experiences related to the war and to family violence were added up to create an indicator for the amount of traumatic experiences due to manmade violence. Figures 2a to 2c show the three different risk gradients for loss of family members, experiences of violence (war and family related), and tsunami severity. Each risk gradient demonstrates a linear negative concurrent (or dose–response) relation between risk and functioning: Exposure to more traumatic events is associated with decrements in positive adaptation.

To examine whether multiplicity of risk stands on its own as a predictor for children's adaptation, even when severity for the different risk types is taken into account, we calculated the same linear regression model for DataSet 4 except that we introduced Multiplicity Index as additional potential predictor. The result of this regression analysis is presented in Table 5.

Results of the regression indicate that the inclusion of a Multiplicity Index neither led to an increase in the variance in children's adaptation explained by the model nor significantly changed the predictive power of the variables related to the severity of single-trauma types. Instead, zero-order correlations show different pictures with both the multiplicity of risk ($r = -.21$) and severity of single trauma types (tsunami severity: $r = -.33$; severity of war exposure: $r = -.25$; amount of family violence: $r = -.27$), being significantly related to children's global adaptation toward risk.

Discussion

The data reported earlier indicate extensive exposure to general adversity and traumatic events among school-aged children in Sri Lanka, specifically Tamil children. Over the three studies that measured parental loss, more than 10% of the 8- to 16-year-old children sampled had already lost a parent. This compares with an average of 7% of the school-age population in Asia (UNAIDS, UNICEF, & USAID, 2004). Socioeconomically, these children were also far more impoverished than their developed world counterparts, with an average of 19 meals per week.

The posttsunami samples indicated that children had extensive exposure to the tidal wave, with 71%–88% of children directly affected by it (being caught or having to flee the wave, or watching people struggling or dying). Among the Tamil children, war exposure was also extremely prevalent, with 59%–89% having experienced at least one war-related life event. Frequently reported war experiences included exposure to bombings and shelling, being close to gunfire or seeing a dead body in the streets.

Relatively limited data were collected about children's experiences of family violence. However, we found an important amount of reported exposure

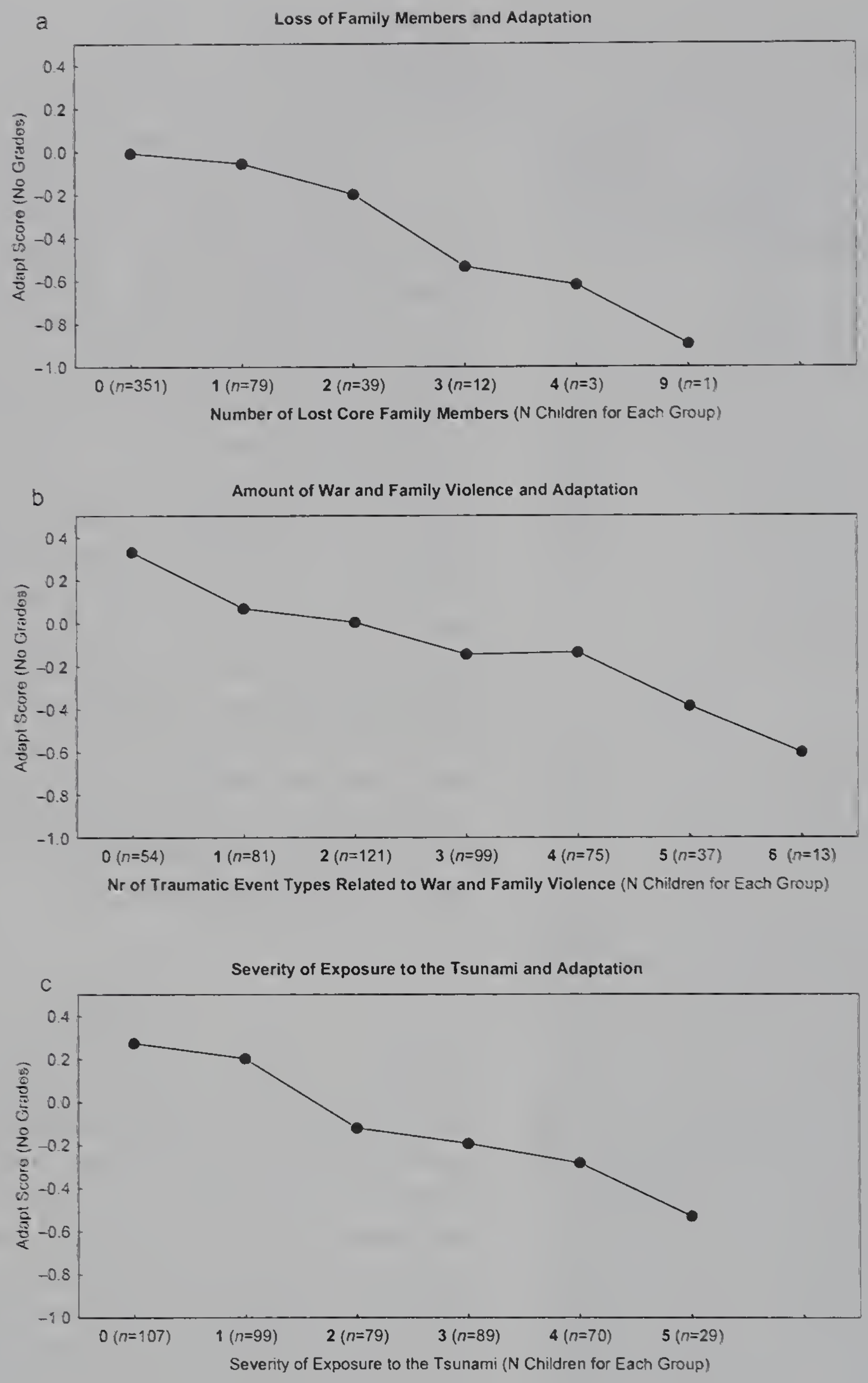


Figure 2. Risk gradients showing positive adaptation outcome for children of the two posttsunami study (Surveys 3 and 4) grouped by the (a) number of lost core family members (top), (b) number of traumatic event types related to man-made (family and war) violence (middle), and (c) severity of tsunami exposure (bottom).

to family violence across the different studies. In Studies 2 and 3, almost one fourth of children (24%–25%) reported being the victim of or witness to a potentially traumatic event related to family violence. In Study 4, where children where asked about a wide range of aversive to stressful to traumatic event types out of the family violence spectrum, 96% of the children reported to have experienced at least one such event with an average of more than five family violence event types for the whole sample. Comparing these data with the numbers of maltreatment at home reported for industrialized countries, the authors concluded that the level of family violence found in the postdisaster Tamil sample can be regarded as exceptionally high (Catani, Jacob, et al., 2008). Being faced with the frequent reports of family violence across the studies, the need for future studies that not only examine stressful and traumatic experiences that happen to children at the individual and community levels but that also include aversive influences occurring within the family becomes particularly salient to us. That would allow us to assess and capture a more holistic picture of children's adjustment (and resilience) in the aftermath of trauma, particularly in environments where the child's

Table 5

Sample 4: Predictive Power of Severity of Different Risk and Trauma Types and Multiplicity of Risks in a Linear Regression Analysis on "Adaptation Score"

Predictor	β	*p* value	Zero-order correlation
Age	−.01	.85	−.04
Gender (male)	.06	.24	−.08
Economic status	.19	< .001	.32*
No. lost family members	−.07	.15	−.09
Amount of family violence event types	−.23	< .0001	−.27*
Amount of war event types	−.13	< .05	−.25*
Tsunami severity	−.32	< .0001	−.33*
Multiplicity Index (number of risk categories)	.001	.98	−.21*

Note. $R^2 = .25$; $F(7, 286) = 13.43$, $p < .0001$.
$^*p < .05$.

bioecological context is threatened at several system levels.

Yet we are keenly aware of the challenges inherent in gathering multilevel data posttraumatic event. Even in developed nations, in addition to the lack of family-level information, relatively few data have been gathered about the details of children's trauma exposure, particularly in the acute aftermath of a disaster or other trauma events (e.g., Gewirtz & Medhanie, 2008; Walsh, 2007). In developing countries, where children's exposure to multiple stressors is much more common—whether as a result of disasters, war, or other forms of organized violence—further information is needed on the nature and extent of trauma in children's lives across systems levels, and its developmental impact.

Indeed, the extent of adversity and trauma exposure among these samples of children provides ample justification for further investigation of the impact of both individual and accumulated traumatic events on children's adjustment.

We are aware that a direct comparison between the explanatory powers of the regression analyses in the different studies is difficult given the variation in assessment and predictor variables. However, the fact that the additional inclusion of a variety of risk types in Studies 2–4 went along with a considerable increase of explained variance points to the importance of incorporating detailed assessment of prior and current traumatic stressors in addition to the stressor of interest (e.g., family violence). Our data are consistent with those in the field of developmental psychopathology (e.g., Cicchetti & Cohen, 1995, 2006; Wachs, 2000) that contend that no single factor is sufficient in explaining developmental outcomes. The risk gradients, depicted in Figures 2a to 2c, highlight the importance of considering the quantity of risk in children's adaptation in the developing world. Although we could find no studies of traumatic events per se that used this framework, our findings are similar to prior studies of stressful life events in the broader developmental literature that have indicated a relation between the number of chronic family stressors and children's adjustment (e.g., McLoyd, 1998; Shaw & Emery, 1988). In addition, however, our data also demonstrate the utility of an overall "adaptation" outcome variable that combines teacher and child report of academic, social, health, and mental health functioning. We have used the term *positive adaptation* rather than *resilience* to reflect our conceptualization of resilience as a *process* of positive adaptation (e.g., Luthar, Cicchetti, & Becker, 2000), to be captured using longitudinal data, rather than the current cross-sectional data. Follow-up data gathered over several subsequent time periods would allow for a consideration of resilience among these at-risk children in Sri Lanka.

A key advantage in modeling both cumulative risk (i.e., different trauma events, as well as life events and socioeconomic risk) and general adaptation is that, as previously discussed, such measurement allows for a broader consideration of the child's ecological context. Beyond the contribution of overall risk to adaptation, the influence of different stressors on adaptation across all samples provides some information about the power of each to disrupt children's lives. The impact of natural disasters, and the tsunami in particular, was evident throughout the studies in which exposure was assessed. Our findings of a powerful impact of the tsunami on functioning a full year after the disaster may reflect the enormous death and destruction that occurred as a result of the tidal wave, beyond the magnitude of many natural disasters (and even wars) in recent history. It may also reflect the impact of a devastating natural disaster overlaid on decades of war and civil unrest. Experiencing war violence significantly influenced adaptation across all samples, regardless of the recency of the tsunami in Study 3. (In Study 2, although the item "being in a place where war was around you" did not significantly predict adaptation, war-related proximal traumatic events did—i.e., "being beaten or shot at" and "witnessing physical violence"). Further, in two of the three studies in which it was measured (Studies 2 and 4), family violence was a highly significant predictor of adaptation. Interestingly, however, in the immediate posttsunami sample (Study 3), loss of family members eclipsed the degree of family violence experienced, in the significance of its association with positive adaptation. Further research is needed to understand predictors of adaptation in the close aftermath of a disaster, but the implication that loss of family members becomes more salient and important to children at this stage following a disaster is consistent with prior observations of ''secure base'' behavior following crises (see Pine, Costello, & Masten, 2005).

Overall, our data sets indicated that all three event types—tsunami and disaster, war, and family violence—significantly contributed to poorer child adaptation. Discrepancies between zero-order correlations and multiple correlation coefficients in the regression analyses indicate that there were suppression effects. However, given the similarity of predictors for children's adaptation found across the data sets, our data show the importance of examining the degree to which severity of exposure as well as moderating factors (within the child and context) might be more powerful contributors to adjustment than the type of trauma event. In fact, as noted earlier, the results of our analysis of both multiplicity and extent and severity of trauma events (Table 5) indicated that multiplicity (i.e., exposure to a greater number of different trauma event types) did not add unique variance to children's adaptation after accounting for severity and extent of exposure to specific trauma types (i.e., war, family violence, tsunami). Although these findings are only applicable to our sample—a sample of children in a high-risk developing country context—they do suggest the importance of further examination of how severity and extent of exposure to a specific risk class versus multiplicity (i.e., exposure to different classes of risk factors) contribute to children's adaptation.

In Studies 1 and 2, child age was also a significant contributor to positive adaptation, with younger age associated with better adaptation, even accounting for the number of trauma event types experienced. In the literature to date, findings regarding the predictive power of age for psychopathology following traumatic events are inconclusive. Whereas some studies have reported younger children to be most vulnerable for the development of trauma-related symptoms (for a discussion, see Dyregrov, Gjestad, & Raundalen, 2002), other studies have yielded different results. For example, a study of children in tsunami-affected areas in Thailand found that older age was significantly associated with depression symptoms (Thienkrua et al., 2006). In particular, studies of trauma-related mental health problems repeatedly have shown that older children are at higher risk of developing PTSD symptoms (Hadi & Llabre, 1998; Khamis, 2005; Nader, Pynoos, Fairbanks, al-Ajeel, & al-Asfour, 1993). With respect to the present data, it should be noted that the number of different event types was assessed rather than the frequency of experiencing each traumatic event. Thus, simply by virtue of their age, older children likely experienced more events within each category. Additionally, older children experienced more of the insecurity related to the civil war that has continued in Sri Lanka for over two decades.

Categorizing data across studies, it appears that in the studies that measured an array of stressors, the independent variables most significantly (negatively) associated with adaptation fall into two categories: (a) amount of exposure to trauma events and (b) loss of family support (either through death or violence). Although severity of the traumatic exposure often has been shown to be associated with compromised outcomes (e.g., Norris et al., 2002), several authors have pointed to the importance of factors within the child and his or her context (e.g., Masten & Coatsworth, 1998; Pine et al., 2005). Examining the contribution of each of these factors (i.e., event-related, or moderating factors—i.e., child and context factors) provides critical questions for further research. Pine et al. (2005) have suggested quantifying the impact of traumatic events on children according to (a) severity of exposure, (b) available family support during and following the event, (c) amount of life disruption engendered by the event, and (d) amount of social disorganization in the child's environment.

Not surprisingly, these categories also have a direct correspondence to children's bioecological micro–macro levels and additionally lend themselves to an evaluation of multiple risk and adaptation or resilience contexts. An implied category provided earlier is the importance of parent–child positive involvement and adaptive parenting practices in the aftermath of trauma (Gewitz, Forgatch, & Wieling, 2008). For example, Gewirtz et al. (2008) argue that the quality of parenting practices following mass trauma might be a key mediator in the relation between trauma exposure and child adjustment. In sum, our data hint at the importance of multiple contextual, past, and current factors in influencing children's adaptation, and point to the need for developmentally and ecologically informed longitudinal studies of children's adaptation following trauma.

Conclusions, Limitations, and Further Research

In their review of disaster research, Norris et al. (2002) report that the extant body of evidence suggests that several populations are particularly vulnerable to the impact of traumatic events: youth, survivors in developing countries, and survivors of mass violence (rather than natural disasters per se). The need for further research to understand the context of trauma in the lives of children in developing countries is clearly evident. As we have demonstrated, such data may provide a unique potential opportunity to assess the additive and interactional contributions of disasters, war, family violence, and other risks on children's development and resilience. For example, Rutter (2005) has highlighted the importance of examining the influence of both independent and nonindependent environmentally mediated risks on development. Disaster events differ from many violence-related events because they are usually nonindependent, although following the disaster event, secondary adversities emerge.

Epidemiological studies, such as the ones reported here, provide a first step in understanding the nature of exposure to traumatic events and adversity on children's adaptation in the developing world. Providing representative sample sizes and population-based data, these studies provide a broad baseline ''snapshot'' of the impact of a variety of traumatic stressors on children's adaptation. In this study, most of the outcome measures used to define positive adaptation refer to the absence of physical or psychological pathology. This is in agreement with Luthar et al. (2000) who have noted: ''The optimal outcome indicators are those that are conceptually most relevant to the risk encountered, so that when there are serious life adversities such as exposure to war, the absence of psychiatric distress can be a more logical outcome than excellence in functioning at school'' (p. 858). Nevertheless, future studies could use broader multi-informant, multimethod measures of adaptation that do not rely solely on the absence of pathology but that incorporate positive, developmentally appropriate domains, such as social competence (e.g., Masten, 2001; Masten et al., 1999).

Moreover, the use of multi-informant data provides added external validity (Rutter, 2005). In situations in which political instability or other factors preclude or limit the capacity for ongoing research, cross-sectional, epidemiological data provide value in understanding the context of risk and protection, and external validity can be maximized by using more than one informant source. The major limitation of this study, however, is its lack of longitudinal data, which precludes the possibility of understanding and disentangling either the sequential effects of different traumatic experiences or the unfolding of their developmental impact.

Pynoos et al. (1999) have proposed a developmental psychopathology model of traumatic stress that places traumatic exposure in the emerging developmental context. As children age, new experiences create opportunities for traumatic reminders, secondary adversities, and new methods for positive adaptation. Further traumatic experiences create potential vulnerabilities that may erode or compromise prior adaptation. In addition to the common observation that age factors intersect with traumatic exposure (i.e., older children have experienced more trauma), developmental factors also interact with factors in the ecological context (e.g., family, school, community, etc.) to enhance protection or vulnerability to adaptation. In addition, studies that have gathered data on different types of traumatic exposure—within and beyond the family—have indicated the interdependence of different traumatic experiences (Landau, 2007; Lynch & Cicchetti, 1998; Walsh, 2007).

Many research questions remain to be answered regarding the impact of trauma on children in the developing world, with its higher base rate of adverse and traumatic events. A developmental risk and protection model framed within a bioecological framework provides a unique opportunity to investigate children's lives in such contexts and might provide invaluable insights into the intersecting role and impact of various system levels in children's overall adjustment posttrauma. Finally, the special value of this research lies in its potential to build a foundation for the development and evaluation of cross-contextual interventions to enhance protective processes for traumatized children in developing countries.

References

Berk, L. E. (2000). *Child development* (5th ed.). Boston: Allyn & Bacon.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Catani, C., Jacob, N., Schauer, E., Mahendran, K., & Neuner, F. (2008). Family violence, war, and natural disasters: A study of the effect of extreme stress on children's mental health in Sri Lanka. *BMC Psychiatry, 8*, 33.

Catani, C., Schauer, E., & Neuner, F. (2008). Beyond individual war trauma: Domestic violence against children in Afghanistan and Sri Lanka. *Journal of Marital and Family Therapy, 34*, 165–176.

Catani, C., Schauer, E., Elbert, T., Missmahl, I., Bette, J.-P. & Neuner, F. (2009). War trauma, child labor, and family violence: Life adversities and PTSD in a sample of school children in Kabul. *Journal of Traumatic Stress, 22*, 162–171.

Catani, C., Schauer, E., Onyut, P. L., Schneider, C., Neuner, F., & Hirth, M. (2005, May). *Prevalence of PTSD and building-block effect in school children of Sri Lanka's northeastern conflict areas*. Paper presented at the annual meeting of the European Society for Traumatic Stress Studies, Stockholm, Sweden.

Cicchetti, D., & Cohen, D. J. (1995). Perspectives on developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 1. Theory and methods* (pp. 3–22). New York: Wiley.

Cicchetti, D., & Cohen, D. (Eds.). (2006). *Developmental psychopathology: Theory and method*. New York: Wiley.

Dyregrov, A., Gjestad, R., & Raundalen, M. (2002). Children exposed to warfare: A longitudinal study. *Journal of Traumatic Stress, 15*, 59–68.

Edleson, J. L. (1999). The overlap between child maltreatment and woman battering. *Violence Against Women, 5*, 134–154.

Elbert, T., Schauer, M., Schauer, E., Huschka, B., Hirth, M., & Neuner, F. (2008). Trauma-related impairment in children—A survey in Sri Lankan provinces affected by armed conflict. *Child Abuse and Neglect, 33*, 238–246.

Garmezy, N. (1993). Children in poverty: Resilience despite risk. *Psychiatry, 56*, 127–136.

Gewitz, A. H., Forgatch, M., & Wieling, E. (2008). Parenting practices as potential mechanisms for children's adjustment following mass trauma. *Journal of Marital and Family Therapy, 34*, 177–192.

Gewirtz, A. H., & Medhanie, A. (2008). Risks to children's adjustment in the aftermath of an incident of domestic violence. *Journal of Emotional Abuse, 8*, 67–83.

Gutman, L. M., Sameroff, A. S., & Eccles, J. S. (2002). The academic achievement of African-American students during early adolescence: An examination of multiple risk, promotive, and protective factors. *American Journal of Community Psychology, 39*, 367–399.

Hadi, F. A., & Llabre, M. M. (1998). The Gulf crisis experience of Kuwaiti children: Psychological and cognitive factors. *Journal of Traumatic Stress, 11*, 45–56.

Hubbard, J., Realmuto, G. M., Northwood, A. J., & Masten, A. S. (1995). Comorbidity of psychiatric diagnoses with posttraumatic stress disorder in survivors of childhood trauma. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*, 1167–1173.

Khamis, V. (2005). Post-traumatic stress disorder among school age Palestinian children. *Child Abuse and Neglect, 29*, 81–95.

Kraemer, H., Kazdin, A., Offord, D., Kessler, R., Jensen, P., & Kupfer, D. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337–343.

La Greca, A., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (1996). Symptoms of posttraumatic stress in children after Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology, 64*, 712–723.

Landau, J. (2007). Enhancing resilience: Families and communities as agents for change. *Family Process, 46*, 351–365.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543–562.

Lynch, M., & Cicchetti, D. (1998). An ecological-transactional analysis of children and contexts: The longitudinal interplay among child maltreatment, community violence, and children's symptomatology. *Development and Psychopathology, 10*, 235–257.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*, 227–238.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from successful children. *American Psychologist, 53*, 205–220.

Masten, A., & Gewirtz, A. (2006). Vulnerability and resilience. In D. Philips & K. McCartney (Eds.), *Blackwell handbook of early childhood development* (pp. 22–43). Oxford, England: Blackwell.

Masten, A. S., Hubbard, J. J., Gest, S. D., Tellegen, A., Garmezy, N., & Ramirez, M. (1999). Competence in the context of adversity: Pathways to resilience and maladaptation from childhood to late adolescence. *Development and Psychopathology, 11*, 143–169.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204.

Nader, K. O., Pynoos, R. S., Fairbanks, L. A., al-Ajeel, M., & al-Asfour, A. (1993). A preliminary study of PTSD and grief among the children of Kuwait following the Gulf crisis. *British Journal of Clinical Psychology, 32*, 407–416.

Neuner, F., Schauer, E., Catani, C., Ruf, M., & Elbert, T. (2006). Post-tsunami stress: A study of posttraumatic stress disorder in children living in three severely affected regions in Sri Lanka. *Journal of Traumatic Stress, 19*, 339–347.

Neuner, F., Schauer, M., Karunakara, U., Klaschik, C., Robert, C., & Elbert, T. (2004). Psychological trauma and evidence for enhanced vulnerability for posttraumatic stress disorder through previous trauma among West Nile refugees. *BMC Psychiatry, 4*, 34.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. (2002). 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry, 65*, 207–239.

Pine, D. S., Costello, J., & Masten, A. (2005). Trauma, proximity, and developmental psychopathology: The effects of war and terrorism on children. *Neuropsychopharmacology, 30*, 1781–1792.

Pynoos, R., Rodriguez, N., Steinberg, A., Stuber, M. L., & Frederick, C. (1998). *UCLA PTSD Index for DSM-IV*. Unpublished manuscript, University of California at Los Angeles.

Pynoos, R. S., Steinberg, A. M., & Piacentini, J. C. (1999). A developmental psychopathology model of childhood traumatic stress and intersection with anxiety disorders. *Biological Psychiatry, 46*, 1542–1554.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M. W. Kent & J. E. Rolf (Eds.), *Primary prevention of psychopathology: Vol. 3. Social competence in children* (pp. 49–74). Hanover, NH: University Press of New England.

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry, 57*, 316–331.

Rutter, M. (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 3–18.

Sameroff, A., Seifer, R., Zax, M., & Barocas, R. (1987). Early indicators of developmental risk: Rochester longitudinal study. *Schizophrenia Bulletin, 13*, 383–394.

Schauer, E., Kohiladevy, M., Catani, C., Onyut, P., Schauer, M., & Neuner, F. (2005, May). *Building local capacity for mental health service provision in the face of large-scale traumatisation: A cascade-model from Sri Lanka*. Paper presented at the annual meeting of the European Society of Traumatic Stress Studies, Stockholm, Sweden.

Shaw, D., & Emery, R. (1988). Chronic family adversity and school-age children's adjustment. *Journal of the American Academy of Child & Adolescent Psychiatry, 27*, 200–206.

Thienkrua, W., Cardozo, B. L., Chakkraband, M. L., Guadamuz, T. E., Pengjuntr, W., & Tantipiwatanaskul, P. (2006). Symptoms of posttraumatic stress disorder and depression among children in tsunami-affected areas in southern Thailand. *Journal of the American Medical Association, 296*, 549–559.

UNAIDS, UNICEF, & USAID. (2004). *Children on the Brink 2004: A joint report of new orphan estimates and a framework for action*. New York: United Nations.

Wachs, T. D. (2000). *Necessary but not sufficient: The respective roles of single and multiple influences on individual development*. Washington, DC: American Psychological Association.

Walsh, F. (2007). Traumatic loss and major disasters: Strengthening family and community resilience. *Family Process, 46*, 207–227.

Werner, E. E., & Smith, R. S. (2001). *Journeys from childhood to midlife: Risk, resilience and recovery*. Ithaca, NY: Cornell University Press.

World Bank. (2006). *Hazards of nature, risks to development*. Retrieved May 9, 2010, from http://siteresources.worldbank.org/INTEVAOFWBASSND/Resources/natural_disasters_evaluation.pdf

Child Development, July/August 2010, Volume 81, Number 4, Pages 1192–1210

Growing Pains: The Impact of Disaster-Related and Daily Stressors on the Psychological and Psychosocial Functioning of Youth in Sri Lanka

Gaithri A. Fernando
California State University, Los Angeles

Kenneth E. Miller
Harvard School of Public Health

Dale E. Berger
Claremont Graduate University

Daily stressors may mediate the relation between exposure to disaster-related stressors and psychological and psychosocial distress among youth in disaster-affected countries. A sample of 427 Sri Lankan Sinhalese, Tamil, and Muslim youth (mean age = 14.5) completed a survey with measures of exposure to disaster-related stressors and daily stressors, psychological distress (posttraumatic stress, depression, and anxiety), and psychosocial distress. The results indicated that daily stressors significantly mediated relations between war- and tsunami-related stressors and psychological and psychosocial distress. Some daily stressors not directly related to disaster also predicted functioning. These results point to the need for policies and interventions that focus on reducing proximal daily stressors that are salient to Sri Lankan youth exposed to disasters.

It is by now a basic tenet of developmental psychology, supported by nearly four decades of research, that children's development is influenced by numerous factors at multiple levels, ranging from intrapersonal variables, both biological and psychological, to factors operating at the level of the family, community, and society (Bronfenbrenner & Morris, 1998; Cooper & Guthrie, 2007). This ecological or multilevel framework, though originally applied to the study of typical or healthy development, is also used to examine psychopathology as well as resilience of youth living in high-stress environments (Buckner, Bassuk, & Beardslee, 2004; Fitzgerald, Karraker, & Luster, 2002; Masten, 2006; Pynoos, Steinberg, & Piacentini, 1999). In their efforts to understand the impact of stressful events and contexts such as parental divorce and poverty on children's emotional, cognitive, and social well-being, researchers have increasingly transcended the earlier emphasis on simple main effects models, looking instead to identify and understand mediating and moderating factors at multiple levels of influence that might explain or influence the relation between stressors and children's mental health (Cicchetti & Lynch, 1993; Osofsky, 2004; Pynoos et al., 1999). The adoption of an ecological framework and the development of risk or causal models that specify mediating and moderating variables at multiple levels of children's social and material ecologies have, in turn, facilitated the development of empirically based interventions that target those variables most amenable to change.

Research on the physical and mental health of children and youth living in settings of armed conflict and natural disaster, or what the Interagency Standing Committee (IASC) has termed *emergency settings* (IASC, 2007), has lagged somewhat in this shift toward an ecological perspective. This lag appears to be related to two major issues. First is the issue of which predictors are, or ought to be, the focus of study. Second is the issue of which outcomes are, or ought to be, the focus of measurement.

This study was made possible by funding from Pomona College, a grant from the Minority Biomedical Research Support program 1R25 GM61331, at California State University, Los Angeles, and by material support from the Centre for Psychosocial Care, Sri Lanka. The authors wish to thank Mr. Jagath Ranawake and the counselors at the Centre for Psychosocial Care and community leaders, teachers, and parents in Ampara who helped with this study. Thanks also to Paul Bolton, who provided important feedback on an earlier version of this article.

Correspondence concerning this article should be addressed to Gaithri A. Fernando, Department of Psychology, California State University, Los Angeles, 5151 State University Dr., Los Angeles, CA 90032. Electronic mail may be sent to gfernan2@calstatela.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0012

Issues Related to Children and Youth Exposed to Disasters

Several reviews synthesize the findings on predictors of mental health in children exposed to situations of armed conflict (Barenbaum, Ruchkin, & Schwab-Stone, 2004; Belfer, 2006; Fazel, Wheeler, & Danesh, 2005; Kalksma-Van Lith, 2007; Stichick, 2001). The reviews demonstrate that the primary focus remains the documentation of the simple or direct relation between exposure to experiences of violence and loss (e.g., exposure to gunfire and bombs, the violent death of family members) and children's mental health. The underlying assumption is that direct exposure to potentially traumatic disaster-related events represents the major threat to children's mental health in emergency settings. Consistent with that assumption, numerous studies have documented a positive—though highly variable—correlation between severity of exposure to war and other disaster-related stressors and psychiatric symptomatology in children, particularly symptoms of posttraumatic stress disorder (PTSD; American Psychiatric Association, 2000; see, e.g., Fazel et al., 2005; Macksoud & Aber, 1996; Neuner, Schauer, Catani, Ruf, & Elbert, 2006; Qouta, Punamäki, & El Sarraj, 2003; Thabet & Vostanis, 2000). This emphasis on the traumatic impact of direct exposure to war- and disaster-related stressors has led to numerous calls for trauma-focused interventions aimed at ameliorating symptoms of PTSD in children living in emergency settings (e.g., Neuner et al., 2006; Yule, 2002).

There is, however, increasing interest in more contextualized approaches that consider multilevel influences on the relation between disasters and children's mental health (Cairns & Dawes, 1996; Dawes & Donald, 1994; Garbarino & Kostelny, 1996; Hart, Galappatti, Boyden, & Armstrong, 2007; Heptinstall, Sethna, & Taylor, 2004; Miller & Rasco, 2004). A growing body of evidence suggests that stressful social and material conditions caused or exacerbated by armed conflict and natural disasters may predict mental health status as well as, or better than, actual degree of exposure to war or to other disaster-related events (Al-Krenawi, Lev-Wiesel, & Sehwail, 2007; Miller, Omidian, Rasmussen, Yaqubi, & Daudzai, 2008; Wessells & Monteiro, 2006). War and natural disasters do a great deal more than expose young people to potentially traumatic stressors; they also generate a host of enduring, stressful conditions of daily life. For the purposes of this study we use the term *daily stressors* to refer to these stressful conditions that may impinge upon children's psychosocial well-being. Such conditions can arise from any area of a child's ecology and include poverty caused or worsened by the loss of family livelihoods, displacement into crowded and impoverished refugee camps with a corresponding loss of privacy and autonomy, a loss of social support caused by the weakening of social networks through the death of family members, alterations in family structure or functioning when family members are killed or disabled, and a lack of access to basic goods and services such as health care and schooling (Barenbaum et al., 2004; Kalksma-Van Lith, 2007; Miller & Rasco, 2004; Wessells & Monteiro, 2006).

Although daily stressors such as those listed earlier may function to mediate or explain the impact of war and other disasters on children's mental health, it seems reasonable to assume that even in situations of complex emergency, some children will also be affected by stressors that may also occur in nonemergency settings, such as family violence and sexual abuse. To the extent that such stressors are unrelated to conditions of war or disaster, they essentially function as independent sources of daily stress in children's lives. This raises the important possibility that, even in emergency settings, at least some of the variance in psychological functioning (including such disaster-related psychiatric conditions as PTSD) may be due to sources of traumatic stress unrelated to conflict or disaster. Thus, one purpose of this study is to examine the relations between direct exposure to war- and tsunami-related stressors, exposure to daily stressors, and distress in children exposed to such stressors.

Issues Relating to Assessment of Outcomes in Children Exposed to Disasters

There is robust research indicating that reactions to trauma in children look qualitatively different from adult reactions to traumatic events (Davis & Siegel, 2005; Osofsky, 2004; Shaw, 2000). However, although researchers acknowledge the developmental and age differences in how children may express trauma reactions, less attention has been paid to cultural differences in such processes. Most studies of children's distress and well-being in complex emergencies have conceptualized these constructs using Western psychiatric diagnoses and self-report measures developed and standardized on Western populations. The most common outcomes examined are depression, anxiety, and child PTSD (for reviews and critiques of this method, see

Barenbaum et al., 2004; Hollifield et al., 2002; Miller, Kulkarni, & Kushner, 2006). Assessments of trauma reactions in children have generally focused on emotional and cognitive symptoms of distress (internal or intrapsychic experiences) while largely overlooking the dimension of psychosocial functioning (internal experiences that affect external, culturally appropriate behaviors). Although a majority of armed conflicts and major natural disasters in the past 50 years have occurred in non-Western nations (UNICEF, 2008; United States Committee for Refugees, 2009), by 2001, only about 14% of samples in disaster research have been from these countries (Norris et al., 2002). When such studies have been conducted, surprisingly little attention has been paid to cultural variations in the ways in which children's well-being and distress are expressed and understood within local cultures.

An understanding of children's functioning that takes social domains into account is particularly salient to collectivistic cultures, where general functioning is dependent on social networks, religious beliefs, cultural practices, and social roles (Summerfield, 1999; Triandis, 1995). Thus, culturally sound assessments of psychosocial functioning merit considerable attention in research conducted with children living in settings of complex emergency. The second goal of the current study was to address this empirical gap. Using a culturally specific measure of children's psychosocial status along with intrapsychic measures of distress, we examined the relations between war and tsunami exposure, daily stressors, and psychological and psychosocial functioning in Sri Lankan youth.

A Contextual Model: Looking Beyond Catastrophic Stressors

Current treatment programs designed to alleviate psychiatric symptoms among children who are exposed to war or natural disasters often focus solely on exposure to the disasters themselves. This emphasis may be overly narrow if in fact, significant variance in children's distress and development is explained by daily stressors caused and exacerbated by, or even unrelated to conflict or natural disaster. Moreover, to the extent that some daily stressors mediate the impact of war or natural disaster on psychosocial functioning and development, it may be especially productive to focus intervention efforts on altering those stressors. War- and other disaster-related traumatic stressors are events that cannot be altered, although their residual impact on psychosocial well-being may be alleviated, but daily stressors represent ongoing, proximal events or situations that continue to adversely affect children, and can be altered.

Few studies have assessed the impact of daily stressors on children's mental health in emergency settings. However, research with adults in zones of ongoing conflict has shown a consistent pattern where the strength of the relation between direct exposure to war and mental health drops considerably when daily stressors are taken into account. Indeed, daily stressors have been demonstrated to account for greater variance in some mental health outcomes, compared with direct exposure to war experiences. For example, among adult refugees from Darfur living in Chadian refugee camps, stressors associated with daily life in the camps such as difficulty accessing basic resources and safety concerns predicted levels of PTSD, as well as indigenous idioms of distress, better than previous exposure to the genocidal violence in Darfur (Rasmussen et al., 2010). Similarly, in a study of adults in Afghanistan, daily stressors such as poverty, unemployment, domestic violence, and social isolation predicted psychological and psychosocial distress better than previous exposure to war-related violence and loss; moreover, daily stressors and war exposure accounted for roughly equal variance in levels of PTSD symptoms (Miller et al., 2009).

In one of the few studies of children that examined the impact of daily stressors in a zone of ongoing conflict, Al-Krenawi et al. (2007) found that family violence was more strongly related to children's mental health than the degree of exposure to political violence. Using ethnographic methods to study families in Afghanistan, de Berry et al. (2003) found that in addition to war-related violence and loss, parents and children identified family violence, poverty, and lack of access to education as major sources of childhood stress. Hart et al. (2007) similarly used a variety of qualitative methods with children in a heavily war-affected region of Sri Lanka, where children identified aspects of armed conflict as well as other threats to their well-being such as poisonous snakebite, theft, and flooding as significant sources of threat to their well-being. Unfortunately, when considering children's mental health in settings of complex emergency, few other studies have looked beyond children's direct exposure to war- or disaster-related traumatic stressors. Consequently, data remain sparse on the relation of daily stressors to the mental health of children living in war- and disaster-affected areas.

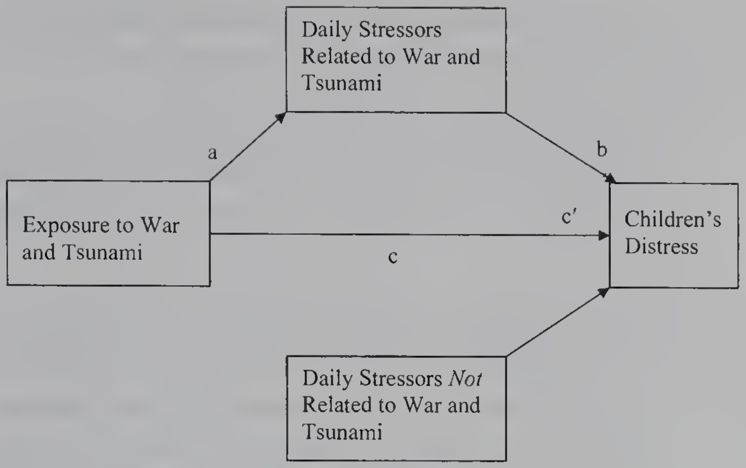

Daily Stressors Related to War and Tsunami
a
b
Exposure to War and Tsunami
c′
c
Children's Distress
Daily Stressors *Not* Related to War and Tsunami

Figure 1. Model for assessing mediation in the relation between disaster exposure, daily stressors, and children's distress.

Figure 1 depicts the hypothesized relation between disasters such as war and daily stressors, and emotional, cognitive, and social distress in youth. In the model, exposure to war and natural disasters exerts a direct effect on distress; however, the influence of this exposure is partially mediated by a subset of daily stressors caused or exacerbated by the conditions of war or natural disaster. There is also a subset of daily stressors that influence children's distress independently of any relation to war and natural disaster. We have intentionally not specified the particular daily stressor(s) in each box (mediators and nonmediators) of the model. Given the current paucity of findings in this area, it seemed premature to specify the relation of particular daily stressors to disasters, on the one hand, and children's distress, on the other. The model in Figure 1 was tested with youth living in a heavily war-torn and tsunami-affected district in eastern Sri Lanka.

Background: Sri Lanka. The island of Sri Lanka has a population of just over 20 million, over 75% of whom are Sinhalese Buddhists. Significant minority ethnic groups include Tamils and Moors, commonly referred to as Muslims (Department of Census and Statistics, Sri Lanka, 2001). Sri Lankans tend to be collectivistic and traditional in orientation (Durvasula & Mylvaganam, 1994; Fernando, 2005, 2008). However, the long-drawn-out armed conflict that began in the 1980s and the tsunami of 2004 appear to have disrupted the social bonds that tie children to parents and families to each other. Somasundaram (2003) has referred to this type of disruption as "collective trauma," a condition where whole communities are so impacted by violence and loss that the very fabric of society is destroyed, exposing people to multiple traumatic events with few safety nets against resulting trauma. Children, being most dependent on their parents and supportive sociocultural structures, are the most likely to be affected by collective trauma. Additionally, children are directly affected by the armed conflict: kidnapped to become child soldiers, being detained and tortured, becoming crippled due to mine explosions, and/or orphaned (Amnesty International, 2009; South Asia Free Media Association [SAFMA], 2008).

A brief respite from the war began in 2002 after a cease-fire agreement was signed by both parties. However on December 24, 2004, just as Sri Lankans were beginning to look forward to rebuilding their lives, the tsunami crashed into Southeast Asia, killing over 35,000 and displacing over a million (SAFMA, 2008). Currently, thousands of families still live in barren, impoverished refugee camps. In addition, the Tamil militants returned to armed conflict in 2005, ironically citing a perceived lack of fairness in the distribution and disbursement of tsunami relief funds as the reason for their return to militancy (SAFMA, 2008). In 2008, over 450,000 people were considered to be living as internally displaced people or refugees due to both the war and the tsunami (SAFMA, 2008).

The Ampara district, where the present study was conducted, is one of the regions most severely affected by the tsunami, with more than 10,000 killed and many more displaced from their homes (Canadian International Development Agency [CIDA], 2009; Department of Census and Statistics, Sri Lanka, 2005). Ampara also represents the eastern frontline of the civil war and was the site of numerous war-related atrocities (Amnesty International, 2008). Many villages, known as "border villages," lie on the border between territories disputed between the Sri Lankan government and the Tamil militants; villagers in these areas often found themselves in the midst of fierce fighting between the two military forces. It is against this background that this study was conducted with Sri Lankan youth.

Issues of age, gender, and ethnic group membership. Few consistent results have been established on the relation between demographic variables (age, gender, and race), disaster exposure, and outcomes for children living in disaster settings. Some studies have found that younger age is a predictor of worse functioning (Bonanno, 2008), whereas others have found no relation between age, exposure, and outcome (for a review and meta-analysis, see Evans, Davies, & DiLillo, 2008). In terms of gender, a robust finding is that girls tend to internalize distress (as depression or anxiety) whereas boys tend to externalize it (as aggressive behaviors) (Evans et al., 2008). A prior study conducted with children in the western and northern parts of Sri Lanka revealed that whereas gender did not significantly predict distress, younger age did (Fernando & Makanui, 2005). The same study found that Muslim and Christian children reported worse psychosocial outcomes, whereas Buddhist and Hindu children reported less distress (Fernando & Makanui, 2005). Because this earlier study was conducted in a different region of Sri Lanka, we will make no predictions in this study as to which demographic variables might be related to predictors and outcomes.

The current study has several goals: (a) to examine the relation of war- and tsunami-related stressors to the psychological and psychosocial functioning of Sri Lankan youth, (b) to examine whether daily stressors mediate this relationship, and (c) to determine which demographic characteristics might be related to distress in Sri Lankan youth. Several hypotheses were tested in this study (see Figure 1); these hypotheses are listed in groups based on the analysis being conducted.

1. War and tsunami exposure will be significantly related to daily stressors (pathway a).
2. War and tsunami exposure will be significantly related to psychological and psychosocial outcomes for Sri Lankan youth (pathway c in Figure 1):
   a. War and tsunami exposure will be significantly related to PTSD, depression, and anxiety.
   b. War and tsunami exposure will be significantly related to overall psychosocial status and to internalizing, externalizing, and psychosocial withdrawal components of psychosocial status.
3. Controlling for the effects of war and tsunami exposure, exposure to daily stressors will be significantly related to psychological and psychosocial outcomes (pathway b):
   a. Exposure to daily stressors will be significantly related to PTSD, depression, and anxiety, after controlling for the effects of war and tsunami exposure; and
   b. After controlling for the effects of war and tsunami exposure, exposure to daily stressors will be significantly related to overall psychosocial status and to internalizing, externalizing, and psychosocial withdrawal components of psychosocial status.
4. Daily stressors will mediate the relation between exposure to the war and tsunami and all outcome measures (pathway c′; weaker than pathway c in Figure 1):
   a. The relation between war and tsunami exposure and the outcomes of PTSD, depression, and anxiety will be weaker after including daily stressors to the model;
   b. Similarly, adding daily stressors to the model will weaken the relation between war and tsunami exposure and overall psychosocial status, and to internalizing, externalizing, and psychosocial withdrawal components of psychosocial status.

If exposure to daily stressors does in fact serve as a mediator in the model, then the regression weight for exposure to war and tsunami predicting outcomes will be smaller when the mediator is also in the model (pathway c′ in Figure 1) than when the mediator is not in the model (pathway c). With full mediation, pathway c′ would be zero. Because no prior studies have considered this model, no prediction was made as to whether daily stressors would fully or partially mediate the relation.

We also examined whether age, gender, and ethnicity are significantly related to disaster exposure, daily stressors, and the outcomes of interest. No hypotheses about these relations were formulated, however, given that prior research has not established such relationships among Sri Lankan youth.

Method

Participants

Participants for this study comprised three groups of students: Sinhalese (n = 223; girls = 108), Tamil (n = 89; girls = 53), and Muslim (n = 115; girls = 79). Student grades ranged from 7 to 13 (Advanced Level II, equivalent to high school senior). Approximately half the sample was female (56%; n = 240). The mean age of the youth was 14.5 years (SD = 1.88), with a range of 11–20 years; no significant differences were found for age by ethnic group. Religious affiliation was exclusively related to ethnicity for the Sinhalese and Moor students (Buddhist and Muslim, respectively). Among the Tamil students, 54 were Hindu whereas 35 were Christian. Most participants' families (97%) could be classified as being of lower socioeconomic status (relative to other Sri Lankans) based on parental income, education, and profession.

Instruments

Demographic survey. The survey included questions about the respondent's gender, age, and education, as well as questions about parental education, income, and profession, type of household (intact, extended families living in the home), type of housing (permanent housing vs. temporary shelter), and religiosity (importance of religious beliefs, frequency of practicing religious beliefs).

The Sri Lankan Index of Psychosocial Status for Children (SLIPSS–C). The SLIPSS–C is a 49-item measure assessing multiple subdomains of psychosocial distress. The measure was developed in Sinhalese and Tamil languages using qualitative methods, including focus groups with children. A more extensive discussion of the development and validation of the measure is available elsewhere (Fernando, Miller, Uyanavithanage, Prasad, & Garcia, 2008). Respondents endorsed items assessing the frequency of feelings and behaviors, from 0 (*never*) to 4 (*almost always*), with higher scores reflecting more psychosocial distress. Internal consistency of the measure was high (Cronbach's alpha = .85) and validity of the scale for Sri Lankan youth has been established (Fernando et al., 2008). Exploratory factor analysis for 46 items using principal axis extraction and direct oblimin (oblique) rotation yielded three factors with eigenvalues > 1.0: internalizing or intrapsychic symptoms (20 items, alpha = .83; e.g., feeling sad, angry, lonely; labeled *S-Internalizing*), externalizing behavior (15 items, alpha = .79; e.g., fighting with friends and neighbors; labeled *S-Externalizing*), and psychosocial withdrawal (11 items, alpha = .65; e.g., difficulty talking to parents and other adults; problems with performing one's rituals and duties; labeled *S-Withdrawal*). The internal consistency of the last subscale is somewhat lower than the others but still considered acceptable (Oyserman, Coon, & Kemmelmeier, 2002). The impact of different types of stressors on the participants' psychosocial status in these three distinct subdomains was examined.

War- and Tsunami-Related Stressor Scale (WTSS). The WTSS is a 12-item scale assessing frequency of exposure from 0 (*never*) to 2 (*more than once*) to various potentially traumatic war-related events (e.g., bomb blasts, kidnappings, shootings) and tsunami-related events (e.g., experiencing the tsunami waves, seeing people die due to the tsunami). Items were selected in consultation with a local team of counselors and based on data from focus groups recruited to develop the SLIPSS–C. Internal consistency (Cronbach's alpha) of the WTSS was .72, although internal consistency may not provide an adequate measure of reliability for this scale (see Netland, 2005). Cronbach's alpha is likely to underestimate reliability for scales where many different experiences are relevant, yet individual items are not necessarily expected to correlate highly. Additionally, although we recognize that any one war- or tsunami-related stressor can have potentially devastating effects, a summative score on the WTSS was used in these analyses. A higher score on the WTSS indicates exposure to multiple war- and tsunami-related events.

Children's Daily Stressor Scale (CDSS). The CDSS includes 20 items generated from focus group transcripts obtained during the development and validation of the SLIPSS–C. The items were based on the most frequent responses to questions posed to focus group participants about stressors they most commonly encountered while living amid poverty, war, and natural disasters. For each item, participants were asked to indicate whether they experienced that event, from 0 (*never*) to 2 (*more than once*). Further details on the development of this measure are provided elsewhere (Miller, Fernando, & Berger, 2009; Miller, Fernando, Uyanavithanage, Prasad, & Ranawake, 2008). Internal consistency (Cronbach's alpha) of the scale was .75. Principal axis extraction and direct oblimin (oblique) rotation revealed three factors, each assessing a subdomain of daily stressors. The first subscale was labeled Deprivation, measured by 10 items (e.g., "Lack of clean drinking water"; alpha = .71). The second subscale was labeled Interparental Conflict, measured by four items (e.g., "Seen your mother or father hitting your father or mother"; alpha = .66). The third subscale was labeled Abuse, measured by six items (e.g., "Been beaten so hard you had injuries"; alpha = .62). The total CDSS score and scores on each of the three subscales were used to examine the contribution of daily stressors to psychosocial status as well as the putative mediating role of these stressors in the relation of war and tsunami exposure to psychosocial status.

Child Posttraumatic Stress Scale (CPSS). The CPSS is a 17-item checklist scale assessing the frequency of symptoms of posttraumatic stress in children exposed to a traumatic event, from 0 (*not at all*) to 3 (*almost always*; Foa, Johnson, Feeny, &

Treadwell, 2001). The CPSS has been used reliably in several studies of children exposed to traumatic stressors (Hawkins & Radcliffe, 2006). Internal consistency of the measure was high (Cronbach's alpha = .81). Respondents were asked to think about the most fearful event they had experienced and to indicate whether, in the past 2 weeks, they experienced any of the feelings, thoughts, or behaviors listed relating to the event they experienced. Higher scores reflect higher levels of PTSD symptomatology.

The Birleson Depression Self-Rating Scale (DSRS). The DSRS is an 18-item scale assessing the frequency of symptoms relating to depression on a scale ranging from 0 (*never*) to 3 (*almost always*; Birleson, Hudson, Grey-Buchanan, & Wolff, 1987). The scale has demonstrated acceptable psychometric properties in studies with children (Myers & Winters, 2002). In the current study, three items were dropped due to poor item–total correlations for this sample, yielding a 15-item scale (Cronbach's alpha = .70).

The Multidimensional Anxiety Scale for Children (MASC). The MASC is a 41-item scale assessing the frequency of symptoms relating to anxiety in children on a frequency scale ranging from 0 (*never*) to 3 (*almost always*; March, Parker, Sullivan, Stallings, & Conners, 1997). A 14-item brief version of the scale was used in this study (Cronbach's alpha = .73).

Note on validity and translation of measures, and terminology The PTSD (CPSS) and depression (DSRS) scales have not been validated for use with Sri Lankan children. Therefore, care was taken during translation and back-translation to ensure equivalency of meaning, comprehensibility, and acceptability to the children (P. Bolton, personal communication, June 5, 2008). The measures were translated into Sinhalese and Tamil by the research team in consultation with the authors. The measures were then back-translated by a Sinhalese psychologist and a Tamil language professor to ensure equivalence of meaning. For ease of communicating results, the CPSS, DSRS, and MASC scales are referred to as PTSD, Depression, and Anxiety scales, respectively, hereafter.

Procedure: Recruitment of Respondents and Administration of Survey

The study was conducted in accordance with the ethical principles detailed by the American Psychological Association and recommendations for disaster research (Kilpatrick, 2004), and was approved by the Institutional Review Board at the first author's institution.

A decision was made by the research team to attempt to recruit youth through the school system because despite the environmental hazards and difficulties, Sri Lankan parents still value education, and ensure that their children attend school regularly. Therefore, school-based samples are representative of the rural population of children in Sri Lanka. Additionally, a decision was made early on to oversample from the ethnic minority groups. Although Sinhalese comprise over 75% of the population of Sri Lanka (Department of Census and Statistics, Sri Lanka, 2001), we oversampled Tamil and Muslim children so that within-group comparisons as well as between-group comparisons could be made. An attempt was made to try to recruit children and youth from late childhood to late adolescence, so that although a broad developmental spectrum could be represented, children would be old enough to understand the questions on the survey without direct intervention by teachers or parents.

The authors worked though a local nongovernmental organization where several counselors represented each ethnic group (Sinhalese, Tamil, and Muslim). Six counselors were trained as research assistants and aided in recruitment, conducting focus groups and discussion groups, writing the scales in comprehensible Sinhalese and Tamil, and administering the surveys. Four schools (two Sinhalese, one Tamil, and one Muslim) were selected for administration of the survey because of their location and the probability of including children in the school who were exposed to the war, the tsunami, or both.

Fully informed written consent was obtained from parents and assent from children prior to administration of the survey. Data were gathered over 4 consecutive days, with the administration of the survey to one school on each day. Children followed along with the survey administrator as each item on the survey was read. All phases of the administration were supervised to assure standardization. The surveys took approximately 1.5 hr (with breaks) to administer to children in Grades 6–8, and less than an hour to administer to children in the upper grades. When the entire administration was completed, children were provided and read a ''Debriefing Form'' and given an opportunity to provide feedback about the experience of having taken the survey. As a token of appreciation, all children were given a package of basic school supplies and a piece of candy, though these had not been promised to the children, to avoid coercion. Finally, each school received a small but needed item identified by the principal, valued at Sri Lankan Rs. 10,000 (approximately US$100).

Of the 537 surveys available for this study, lack of consent forms led to 16 dropped surveys and another 94 were dropped because of questions about the validity of the data (the SLIPSS–C included three validity checks; Fernando et al., 2008). Cases with data missing on any of the variables used in the mediation analyses were also dropped, leaving 427 cases for the current analyses. Analyses of surveys with missing data did not show significant differences in age, gender, or ethnicity between the group remaining in the study and the group dropped from further analyses.

Notes on analyses. Because we conducted numerous analyses, an alpha level of .001 was used as the general criterion for statistical significance, although alpha values of .01 and .05 are shown to reveal patterns in the results. Effect sizes are shown as correlations, as beta weights, as R^2 added in regression analyses, and as group means with standard deviations for group comparisons. We report item means rather than the mean of total scores on the various scales to facilitate interpretations relative to the scale anchors and enable comparisons of scales that use the same anchors but have differing numbers of items (e.g., the CDSS has 20 items and the WTSS has 12 items, but both use 0–2 scale anchors). Nonparametric tests are reported in some cases where distributional assumptions of parametric tests are not met and nonparametric tests fail to support the parametric results. Although the sample size is $n = 427$ for all analyses that use all groups, the degrees of freedom for t tests vary because we used the test that does not assume equal variances.

Results

Analyses

Several analyses were conducted in accordance with the hypotheses and the exploratory examinations mentioned earlier. First, we examined whether age, gender, or ethnicity were significantly related to exposure to disaster-related and daily stressors. Second, we examined whether the outcomes of posttraumatic stress, depression, anxiety, and psychosocial status differed by age, gender, or ethnicity.

Next, we tested each of the hypotheses listed earlier in this study. Correlation and regression analyses were conducted in the following manner:

1. To test our first hypothesis, that exposure to disaster-related stressors was significantly related to daily stressors, a correlational analysis was conducted between WTSS scores and CDSS scores.
2. To test the second set of hypotheses that exposure to the war or tsunami will predict psychological and psychosocial outcomes, simple regressions were conducted with WTSS scores as the only predictor for each of the measured outcomes (PTSD, Depression, Anxiety, SLIPSS–C, S-Internalizing, S-Externalizing, and S-Withdrawal).
3. To test the third set of hypotheses, that exposure to daily stressors in general would predict psychological and psychosocial outcomes after accounting for exposure to disaster-related stressors, total CDSS scores were added as a second step in multiple regression analyses for each of the outcomes.
4. To test the fourth set of hypotheses, that exposure to daily stressors will mediate the relation between exposure to disaster-related stressors and outcomes, the regression weights in each model were examined. If the beta weight for WTSS obtained in the first step of each model is reduced when CDSS is entered into the model, mediation is demonstrated.
    a. To examine whether specific types of daily stressors mediated the relation between disaster-related stressors and outcomes, each of the three subscale scores of the CDSS were entered at the second step of each of the multiple regression models. Because three conceptually distinct factors emerged from factor-analyzing the CDSS, separate mediation analyses were conducted using the Deprivation, Interparental Conflict, and Abuse subscales of the CDSS.
    b. Finally, demographic variables were entered into each of the six models, to assess whether any of these variables would further clarify the relation between the predictors and outcomes. As noted earlier this was an exploratory analysis because there are no robust data on which to build hypotheses on such variables as age, gender, and ethnic group membership in this sample.

The results of these analyses are presented as follows.

Demographic Comparisons

Table 1 provides information on age, gender, and ethnicity for exposure to stressors using the mean of the item scores as the reported statistic. The differences are reported for each group along with an indication of statistical significance ($p < .001$) for pairwise group differences using t tests not assuming equal variances. Examination of the relation between age and stress exposure revealed that age was not significantly related to WTSS scores, $r(427) = .04$, *ns*, but was positively related to CDSS scores, $r(427) = .23$, $p < .001$. This relation was particularly strong for the Deprivation scale of the CDSS, with older children reporting greater deprivation, $r(427) = .33$, $p < .001$. Gender differences were not significant for WTSS or overall CDSS scores.

Examining ethnic group differences in exposure to stressors revealed that Tamil and Muslim youth obtained significantly higher mean WTSS scores (indicating greater exposure to multiple war- and tsunami-related stressors) compared to Sinhalese youth. The three groups did not differ significantly on CDSS (daily stressors) total scores. However, the groups did differ on the subscales of the CDSS. Compared to the Sinhalese youth, Muslim and Tamil youth obtained significantly higher mean Deprivation scores. Muslim youth obtained a significantly higher mean Abuse score but a significantly lower Interparental Conflict score, compared to Sinhalese and Tamil youth.

Table 2 provides information on the relation between demographic variables and outcome measures. Age was not significantly related to any of the outcome measures. Significant gender differences were obtained for two subscales of the SLIPSS–C; boys obtained significantly higher S-Externalizing scores, $t(376) = 4.56$, $p < .001$ ($d' = .45$), whereas girls obtained significantly higher S-Internalizing scores, $t(418) = 3.76$, $p < .001$ ($d' = .36$). Examination of scores obtained by ethnic groups revealed that Tamil and Muslim youth obtained significantly higher mean SLIPSS–C scores compared to Sinhalese youth. Subscale analysis revealed an especially strong ethnic group difference on S-Internalizing, with Muslim youth obtaining the highest mean score, Sinhalese youth obtaining the lowest, and Tamil youth scoring between these two groups (all pairwise differences significant at $p < .001$).

Other outcome measures also revealed significant ethnic group differences. Tamil and Muslim youth obtained significantly higher mean Depression scores compared to Sinhalese youth, and Muslim youth obtained a significantly higher mean Anxiety score compared to Sinhalese and Tamil youth. The ethnic groups did not differ significantly on PTSD. Overall, psychological and psychosocial outcomes appear to be worst for Muslim youth, whereas Sinhalese youth reported the least amount of distress. Tamil youth fell somewhere between these two groups. The relation of the demographic variables of gender, ethnicity, and age to outcome

Table 1
Age, Gender, and Ethnic Group Comparisons on Trauma Exposure (n = 427)

	Trauma exposure variables				
	WTSS	CDSS (Total)	Deprivation	Interparental conflict	Abuse
Number of items	12	29	10	4	6
Cronbach's alpha	.72	.75	.71	.66	.62
Possible range	0–2	0–2	0–2	0–2	0–2
r with age (Pearson r)	.04	.23***	.33***	.15	–.12
Gender, M (SD) of mean scores					
Female ($n = 240$)	0.29_a (0.24)	0.44_a (0.21)	0.62_a (0.37)	0.36_a (0.47)	0.12_a (0.21)
Male ($n = 187$)	0.32_a (0.29)	0.49_a (0.26)	0.62_a (0.40)	0.45_a (0.47)	0.22_b (0.34)
Ethnic group, M (SD) of mean scores					
Sinhalese ($n = 223$)	0.84_a (0.39)	0.43_a (0.21)	0.52_a (0.31)	0.47_a (0.49)	0.13_a (0.26)
Tamil ($n = 89$)	0.92_b (0.42)	0.48_a (0.24)	0.72_b (0.39)	0.39_a (0.45)	0.17_a (0.23)
Muslim ($n = 115$)	1.18_b (0.48)	0.50_a (0.28)	0.75_b (0.45)	0.28_b (0.41)	0.24_b (0.32)

Note. Group means within gender and ethnicity that do not share the same subscript differ at $p < .001$. WTSS = war- and tsunami-related stressor scale; CDSS = Children's Daily Stressor Scale.
***$p < .001$.

Table 2
Age, Gender, and Ethnic Group Comparisons on Outcome Variables (n = 427)

	Outcome variables						
	PTSD	Depression	Anxiety	SLIPSS–C	S-Internalizing	S-Externalizing	S-Withdrawal
Number of items	17	15	14	49	20	15	11
Cronbach's alpha	.81	.70	.73	.85	.83	.79	.65
Possible scale range	0–3	0–3	0–3	0–4	0–4	0–4	0–4
Age (Pearson *r*)	.099	−.027	.085	−.010	.040	−.076	−.047
Gender, *M* (*SD*)							
Female (n = 240)	1.09_a (0.49)	0.78_a (0.33)	1.00_a (0.45)	1.15_a (0.34)	1.47_a (0.51)	0.71_a (0.40)	1.26_a (0.47)
Male (n = 187)	1.06_a (0.45)	0.82_a (0.37)	0.88_a (0.42)	1.17_a (0.34)	1.29_b (0.45)	0.91_b (0.45)	1.36_a (0.47)
Ethnicity, *M* (*SD*)							
Sinhalese (n = 223)	1.04_a (0.48)	0.73_a (0.32)	0.84_a (0.39)	1.06_a (0.29)	1.22_a (0.41)	0.75_a (0.36)	1.26_a (0.48)
Tamil (n = 89)	1.04_a (0.43)	0.87_b (0.37)	0.92_a (0.42)	1.20_b (0.34)	1.46_b (0.46)	0.82_a (0.51)	1.31_a (0.46)
Muslim (n = 115)	1.17_a (0.48)	0.87_b (0.34)	1.18_b (0.48)	1.32_b (0.36)	1.68^c (0.53)	0.88_a (0.49)	1.37_a (0.44)

Note. Group means within gender and ethnicity that do not share the same subscript differ at $p < .001$. PTSD = posttraumatic stress disorder; SLIPSS–C = Sri Lankan Index of Psychosocial Status for Children; S-Internalizing = Internalizing subscale of the SLIPSS–C; S-Externalizing = Externalizing subscale of the SLIPSS–C; S-Withdrawal = Withdrawal subscale of the SLIPSS–C.

measures will be examined in subsequent regression analyses.

The Relation Between War- and Tsunami-Related Stressors and Daily Stressors

A simple correlation analysis indicated a significant relation between WTSS scores and CDSS scores, $r = .44$ ($n = 427$), $p < .001$, confirming the first hypothesis (path a in Figure 1). Children with greater exposure to disasters experienced higher levels of daily stressors.

The Relation Between War- and Tsunami-Related Stressors and Psychological and Psychosocial Functioning

Regression analyses using WTSS as the only predictor for each of the outcomes indicated that WTSS scores were significantly related to each outcome (see Model 1 in Table 3) except the S-Withdrawal subscale of the SLIPSS–C (path c in Figure 1). Exposure to war- and tsunami-related stressors was significantly related to increased PTSD ($\beta = .32$), Depression ($\beta = .23$), and Anxiety ($\beta = .33$), as well as to total SLIPSS–C ($\beta = .35$) scores and to two of its subscales, S-Internalizing ($\beta = .40$) and S-Externalizing ($\beta = .23$) scores ($p < .001$ for each of these outcomes). Because WTSS scores were not significantly related to S-Withdrawal scores ($\beta = .01$), no further regression analyses were conducted with this subscale. Thus, the second set of hypotheses was confirmed for all but S-Withdrawal. Exposure to stressors associated with war and tsunami significantly predicted poorer psychological and psychosocial functioning.

The Contribution of Daily Stressors Beyond WTSS in Predicting Psychological and Psychosocial Functioning

Regression analyses using both WTSS and CDSS as predictors indicated that CDSS scores contributed beyond WTSS to predict all six outcome measures (see Model 2 in Table 3; path b in Figure 1). Controlling for WTSS, the CDSS total score significantly predicted PTSD ($\beta = .40$), Depression ($\beta = .22$), and Anxiety ($\beta = .35$) scores, as well as the overall SLIPSS–C ($\beta = .25$) and the two subscales of S-Internalizing ($\beta = .27$) and S-Externalizing ($\beta = .19$) ($p < .001$ for all analyses) scores. Thus, the third set of hypotheses was confirmed: Exposure to daily stressors was significantly predictive of poorer psychological and psychosocial functioning, beyond that which can be predicted by exposure to the war and tsunami.

Daily Stressors as Mediators Between War- and Tsunami-Related Stressors and Outcomes

Total CDSS scale as a mediator. Model 1 in Table 3 provides the beta weights of the WTSS in regression analyses predicting each of the six outcome variables. These weights are represented by path c in Figure 1. Model 2 in Table 3 provides the regression weights when both WTSS and CDSS are used to predict each of the outcomes. The regression weights for WTSS and CDSS correspond to paths c′ and b, respectively, in Figure 1. The corre-

Table 3
Summary of Regression Analyses Testing Potential Mediating Effects of Daily Stressors on the Relations Between Disaster Exposure and Psychological and Psychosocial Outcomes (n = 427)

			β of predictors on outcome variables					
Model	Predictors	*r* with WTSS	PTSD	Depression	Anxiety	SLIPSS–C	S-Internalizing	S-Externalizing
1	WTSS		**.32**	**.23**	**.33**	**.35**	**.40**	**.23**
2	WTSS		**.14**	.13*	**.17**	**.24**	**.29**	.14**
	CDSS	**.44**	**.40**	**.22**	**.35**	**.25**	**.27**	**.19**
3	WTSS		**.19**	**.18**	**.21**	**.30**	**.33**	**.23**
	Deprivation	**.42**	**.29**	.12*	**.28**	.11*	**.18**	.00
4	WTSS		**.31**	**.23**	**.32**	**.34**	**.40**	**.23**
	Interparental Conflict	.02	**.19**	.14**	.14**	**.17**	.12**	**.15**
5	WTSS		**.25**	**.18**	**.26**	**.27**	**.34**	.14**
	Abuse	**.27**	**.23**	**.20**	**.24**	**.30**	**.23**	**.33**
6	WTSS		**.17**	.15**	**.18**	**.25**	**.28**	**.17**
	Deprivation		**.24**	.08	**.23**	.05	**.15**	−.07
	Interparental Conflict		.11**	.09	.06	.11*	.05	.11*
	Abuse		**.18**	**.17**	**.20**	**.27**	**.21**	**.32**
R^2 added	WTSS		**.10**	**.05**	**.11**	**.12**	**.16**	**.05**
in Model 2	CDSS		**.13**	**.04**	**.10**	**.05**	**.06**	**.03**

Note. All tests are two-tailed. Numbers in bold indicate $p < .001$. PTSD = posttraumatic stress disorder; SLIPSS–C = Sri Lankan Index of Psychosocial Status for Children; S-Internalizing = Internalizing subscale of the SLIPSS–C; S-Externalizing = Externalizing subscale of the SLIPSS–C; WTSS = War- and Tsunami-Related Stressor Scale; CDSS = Children's Daily Stressor Scale.
$^*p < .05$. $^{**}p < .01$.

lation between WTSS and CDSS is .44, represented by path a. The mediation effect can be observed as the reduction in the beta weight for WTSS when the mediator is added to the model. For predicting PTSD, we find this reduction to be substantial: c–c′ = .315–.137 = .178.

An equivalent way to compute the mediation effect is the product of the regression weights for the paths through the mediator, paths a and b. For predicting PTSD with CDSS as a mediator, this product is .444 × .401 = .178, the same value as c–c′. Both paths a and b are statistically significant with $p < .001$, which we take as evidence that daily stressors (CDSS) mediate at least partially the relation between exposure to the stressors of war and tsunami (WTSS) and PTSD. Statistical significance of the a × b index of mediation can also be tested with the Sobel test (Sobel, 1982), which yields $z = 5.33$, $p < .001$.

The beta weights for the CDSS (Model 2 in Table 3) are statistically significant ($p < .001$) for each of the six outcome variables. Thus, the CDSS mediates between WTSS scores and scores on PTSD, Depression, Anxiety, and SLIPSS–C (total, as well as S-Internalizing and S-Externalizing) scales. In each model, the mediation effect is partial rather than complete because although the beta weights on the WTSS are reduced for all outcome measures, they do not reduce to zero but rather remain statistically significant with $p < .001$ for PTSD, SLIPSS–C, S-Internalizing, and Anxiety scores.

The final rows of Table 3 provide the R^2 added when WTSS is entered at the first step of a hierarchical regression and when CDSS entered at the second step ($p < .001$ for all six outcome variables). Examination of these R^2 values demonstrates that the CDSS approximately doubles the variance explained by the WTSS alone in predicting PTSD, Depression, and Anxiety.

The individual subscales of the CDSS as mediators. Because CDSS scores were significantly and meaningfully related to the six outcome measures, and because three conceptually distinct factors emerged from factor analyzing the CDSS, separate mediation analyses were conducted using each of the three subscales of the CDSS (Deprivation, Interparental Conflict, and Abuse; Models 3, 4, and 5 in Table 3).

As a first step, the relations between WTSS scores and each of the three subscale scores (path a) were tested. The WTSS was correlated with two of the three subscale scores of CDSS: Deprivation ($r = .42$, $p < .001$) and Abuse ($r = .27$, $p < .001$). Because path a was not significant for Interparental Conflict ($r = .02$, $p > .05$), this factor cannot function as a mediator between WTSS and the outcome variables; no further mediation analyses were conducted with this variable.

As shown in Model 5 (Table 3), Abuse was a significant predictor of each measure of psychological and psychosocial functioning, beyond the contribution of WTSS. The beta values ranged from .20 for Depression to .33 for S-Externalizing ($p < .001$ for each of these path b beta weights). Thus, Abuse was a strong mediator of the relation between WTSS and each of the outcome measures. Deprivation also clearly mediated the relation between WTSS and three of the outcome measures: PTSD ($\beta = .29$), Anxiety ($\beta = .28$), and S-Internalizing ($\beta = .18$), all $p < .001$. Evidence for mediation was present but weaker for the total SLIPSS–C ($\beta = .11$, $p < .05$) and Depression ($\beta = .12$, $p < .05$) scores. These results confirm our fourth set of hypotheses, providing strong evidence for mediating effects of the daily stressors of deprivation and abuse on the relationship between exposure to war and the tsunami and distress in youth.

Multiple mediator models. The three subscales of the CDSS were moderately correlated. The correlation of Deprivation with Interparental Conflict was $r = .19$ and with Abuse was $r = .23$, whereas the correlation between Interparental Conflict and Abuse was $r = .19$ (all $ns = 427$, $ps < .001$).

Mediation effects and the unique contributions of each variable are shown in Model 6 (Table 3). The Abuse subscale of the CDSS demonstrated significant and unique mediating effects with all outcome measures ($p < .001$ for the relation with each outcome variable). The direct paths (c′) for the WTSS were significantly greater than zero ($p < .001$ for all outcome variables except Depression scores, where $p < .01$), indicating that the mediation effect was only partial. The unique mediation effects of Deprivation and Interparental Conflict scores were generally weaker, although the Deprivation subscale showed unique mediation effects in predicting PTSD, S-Internalizing, and Anxiety scores ($p < .001$).

When age, gender, and ethnicity were included along with the WTSS and the three CDSS subscales as predictors, age did not add unique variance to predicting any of the six psychological and psychosocial outcomes, whereas gender and ethnicity did explain variance in some models. Controlling for all other variables, being female predicted higher Anxiety ($\beta = -.16$) and S-Internalizing ($\beta = -.20$), whereas being male predicted higher S-Externalizing ($\beta = .17$). Compared to the Sinhalese youth, Muslim youth obtained significantly higher Anxiety ($\beta = .17$), total SLIPSS–C ($\beta = .24$), and S-Internalizing ($\beta = .23$) scores. Even with the demographic variables entered into the model, CDSS scores contributed beyond these variables and WTSS scores in predicting all six outcomes, with R^2 changes of .12 for PTSD and .10 for total SLIPSS–C scores. When demographic variables were included in the model, WTSS scores contributed uniquely toward predicting PTSD, SLIPSS–C, and S-Internalizing scores but not to predicting the other three outcome variables. Space limitations preclude more in-depth analyses of the relation of the demographic variables with the outcome measures but are available elsewhere (Fernando et al., 2008).

Summary of Analyses

Taken together, the results convincingly demonstrate that daily stressors mediate the relation between exposure to disaster-related stressors and psychological and psychosocial outcomes for Sri Lankan youth. As summarized in Table 3, the total CDSS score and the subscales of Deprivation and Abuse significantly mediated the relation between WTSS scores and most of the outcomes (PTSD, depression, anxiety, and two subdomains of psychosocial status). These findings held even when age, gender, and ethnicity were statistically controlled. The subscales of daily stressors demonstrated distinct predictive patterns (see Table 3, Model 6) with all predictors in the model: Deprivation was a strong predictor of PTSD and Anxiety but not of Psychosocial Status. In contrast, Abuse was a relatively stronger predictor of Psychosocial Status than of PTSD. In particular, S-Externalizing was strongly predicted by Abuse ($\beta = .32$) but not by Deprivation ($\beta = .07$). Figure 2 uses data from Table 3 (Models 1 and 6) to graphically represent how two of the three subscales of the CDSS mediate the relation between the WTSS and PTSD.

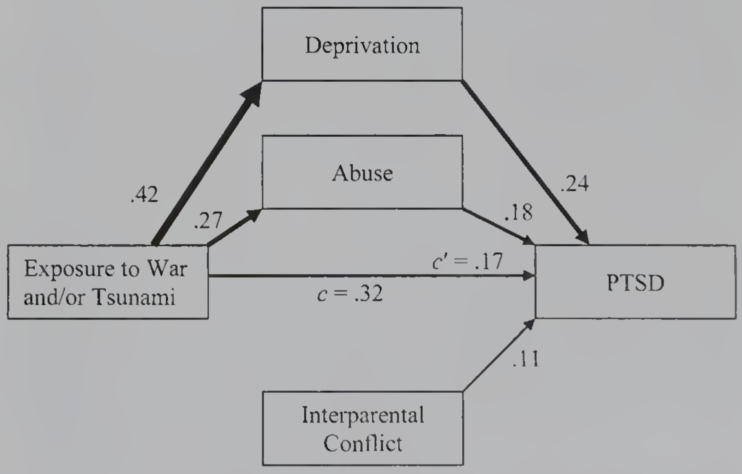


Figure 2. Daily stressors as mediators of the relation between exposure to war and/or tsunami and posttraumatic stress disorder ($n = 427$).

Discussion

In this study, we sought to examine the relative contribution of disaster exposure, on one hand, and daily stressors, on the other, to the mental health and psychosocial functioning of youth in a region of eastern Sri Lanka sorely affected by a prolonged civil war and a devastating tsunami. We also sought to assess the extent to which daily stressors mediate or explain the relation between disaster exposure and children's mental health. Using a multiethnic sample that reflected the ethnically diverse population of Sri Lanka in general and the Ampara district in particular, we used focus groups with youth to identify salient, proximal stressors in their social and material ecologies. Items were used to develop a self-report measure of daily stressors (the CDSS), which was subsequently used with self-report measures of war and tsunami exposure, PTSD, depression, anxiety, and a culturally specific measure of psychosocial distress.

The two issues we highlighted earlier in this study were addressed in this study. First was the tendency of researchers to focus solely on direct exposure to disasters as predictors of distress in children. As in much of the previous research with youth in settings of armed conflict and natural disaster, the results of this study demonstrated that exposure to disaster-related violence and loss was significantly related to the psychological and psychosocial distress of Sri Lankan youth. Limiting our analyses to these relations, a strong case could be made for trauma-focused interventions to be conducted with this group. However, and consistent with our hypotheses, the addition of daily stressors to the models markedly reduced the power of disaster exposure to predict the psychological and psychosocial distress of Sri Lankan youth, even that of PTSD (see Figure 2). The daily stressors taken together were stronger predictors of PTSD than direct exposure to the conflict and the tsunami, and about as strong a predictor of the other measures of psychological and psychosocial distress. Moreover, as hypothesized, daily stressors partially mediated the relation between direct exposure to conflict and tsunami and resulting distress of Sri Lankan youth. Thus, the failure to include daily stressors in explanatory models of distress among young people in emergency settings is likely to overestimate significantly the predictive power of disaster exposure, and to overlook critical sources of variance in youth well-being or distress.

When the CDSS subscales identified through factor analysis were entered into the mediation models, several key findings emerged, all consistent with our hypotheses and the models depicted in Figures 1 and 2. First, material deprivation, witnessing interparental conflict, and being the victim of child abuse all contributed significantly to psychological and psychosocial distress, although the strength of these relations varied depending on the outcome variable used in each model. Importantly, both child abuse and deprivation predicted PTSD as well as or better than direct exposure to the war and tsunami (see Table 3 and Figure 2); clearly, therefore, our findings suggest that *direct exposure to armed conflict and natural disaster was neither the only nor even the primary source of traumatic stress in these children's lives*. We also found, as hypothesized, that the effects of exposure to disaster on the functioning of these Sri Lankan youth were at least partially mediated by experiences of material deprivation and child abuse. This study found empirical evidence that conflict and natural disasters generate conditions of deprivation for children and their families (e.g., diminished access to basic resources). That such deprivation should in turn affect young people's well-being is not surprising.

On the other hand, the mediating role of child abuse is less readily interpreted. The findings, however, do support Somasundaram's (2003) claim that the war, in particular, may be creating a culture of violence, where overall incidents of individual and family violence increase. In addition to physical abuse, several items on the abuse subscale also reflect actual or feared experiences of sexual abuse, and were endorsed by youth of both sexes (with boys reporting more sexual and physical abuse than girls). It may be that the generally stressful social conditions created by disaster reduce the patience and increase the reactivity of parents and teachers, making physical abuse (especially against boys) more likely. It may also be that conditions of war weaken social norms that otherwise protect children from being physically and sexually victimized. Additionally, in the crowded confines of camps for displaced families, privacy and parental protection are diminished and the potential for sexual abuse is increased.

We did not assess the sources of actual or feared abuse; consequently, we do not know whether the youth in our sample were primarily concerned with physical and sexual abuse by teachers, older youth, relatives, armed combatants, or other adults such as police officers. Anecdotally, we have heard that older school boys sometimes prey upon younger boys to obtain sexual pleasure, as sexual contact between young men and women is restricted in Sri Lanka due to strict sociocultural and religious norms. Follow-up research, conducted in a culturally sensitive manner, could examine the sources and nature of sexual and physical abuse of both young men and women in Sri Lanka, and how conditions of war and natural disasters might exacerbate these traumatic, yet preventable events.

Finally, as hypothesized in Figure 1 and depicted as an example in Figure 2, a subset of daily stressors, interparental conflict, was unrelated to war and tsunami exposure but did exert a significant effect on PTSD, anxiety, psychosocial status in general, and externalizing psychosocial behavior in particular. Focusing solely on disaster-related variables would have failed to pick up these important relations between daily stressors and distress among Sri Lankan youth.

The second issue we addressed was the tendency of some trauma researchers to focus on PTSD and other psychiatric symptomatology as the main outcomes of interest when assessing distress in youth living in disaster settings. The results of this study indicate that although PTSD was a valid outcome to assess, the addition of the domain of psychosocial status provided a broader and more culturally relevant understanding of the relation between disaster exposure and distress in Sri Lankan youth. Gender and ethnic group differences were obtained for psychosocial status, particularly for internalizing behavior, where girls and Muslim youth reported significantly higher internalizing behavior compared to boys and to Sinhalese and Tamil youth, respectively. Tamil youth in turn reported significantly higher internalizing behavior compared to Sinhalese youth. No such group differences were obtained for PTSD scores. Given that Muslim youth also reported significantly higher exposure to the war and tsunami, as well as to daily stressors, it is interesting that the outcome of PTSD, which is sensitive to trauma exposure, did not show such group differences. Thus, in this study, the addition of a locally developed measure of psychosocial status provided a more comprehensive picture of how both disaster-related and daily stressors impact the psychosocial well-being of youth, compared with focusing solely on PTSD as an outcome. Additionally, although we did not focus on the outcome of anxiety as a measure of distress, the MASC (the measure of anxiety used for this study) was more sensitive to both disaster-related and daily stressors, compared with the measure of PTSD. This suggests that when working in Sri Lanka with scarce resources, it may be more useful to use the SLIPSS–C and MASC as outcome measures.

The Relation of Demographic Variables to Predictors and Outcomes

Examination of the relation between demographic variables (age, gender, and ethnic group membership) and exposure to disaster-related stressors and daily stressors revealed that although older age was related to greater exposure to daily stressors, age did not significantly predict any assessed outcomes. Gender was not significantly related to predictors but was significantly related to psychosocial outcomes; consistent with other studies (e.g., Montgomery, 2008), girls reported significantly higher internalizing tendencies and anxiety, whereas boys reported significantly higher externalizing behaviors.

Ethnic group membership was related to both predictors and outcomes. Exposure to disaster-related stressors was greater for the Tamil and Muslim youth than for Sinhalese youth. Additionally, Muslim and Tamil youth showed greater levels of deprivation than Sinhalese youth, and Muslim youth reported higher levels of abuse than Sinhalese youth. Conversely, Sinhalese youth reported significantly greater exposure to interparental conflict compared to Muslim youth. In terms of outcomes, the most striking difference was between Sinhalese and Muslim youth; the latter reported significantly higher depression, anxiety, and internalized psychosocial distress, compared to Sinhalese children. There is evidence that Sri Lankan Muslims are often ignored in the debate about where to allocate scarce economic resources (International Crisis Group, 2007); these results speak to an urgent need to ensure that intervention programs target this at-risk ethnic minority group. Additionally, more studies should be conducted to determine whether the findings obtained here about specific ethnic groups enduring different types of hardship are robust.

Implications of Results for Practice and Policy

There is a growing debate regarding the relative priority of trauma-focused versus more holistic psychosocial interventions with children in settings of armed conflict and natural disaster. On one side of this debate, proponents of trauma-focused models point to elevated rates of PTSD symptoms among children in emergency settings as evidence that intervention efforts should aim primarily at the amelioration of psychological trauma resulting from exposure to war and disaster using specialized trauma treatment programs (Neuner et al., 2006; Yule, 2002). Implicit in this view is a particular set of assumptions: (a) PTSD is valid and meaningful diagnostic construct for children regardless of the cultural context; (b) PTSD, rather than other expressions of distress, represents the most critical indicator of children's mental health in situations of war and disaster; and (c) children's distress in emergency settings is primarily related to traumatic exposure to war- and disaster-related violence and loss.

Advocates of more holistic psychosocial approaches challenge these assumptions and argue instead for community-based approaches that strengthen naturally occurring support systems, normalize children's daily routines, and provide child-friendly spaces in schools and other locales for activities that promote recovery without focusing specifically on trauma. These researchers maintain that the cross-cultural validity of PTSD has generally been assumed rather than systematically evaluated, and that even when symptoms of PTSD are present, they may be of secondary concern to families in cultural contexts where other expressions of distress may hold greater salience (de Berry et al., 2003; Hart et al., 2007; Kalksma-Van Lith, 2007; Miller & Fernando, 2008; Miller et al., 2009; Wessells & Monteiro, 2006). These researchers also argue that by focusing so narrowly on the posttraumatic effects of war- and disaster-related violence and loss, insufficient attention is paid to *ongoing* sources of stress in children's lives that may be amenable to change through targeted interventions (Barenbaum et al., 2004; Belfer, 2006). Finally, advocates of psychosocial approaches point to findings that symptoms of distress, including trauma, tend to abate in many children given adequate social support, the normalization of circumstances, and the passing of time (Belfer, 2006; Thabet & Vostanis, 2000; Wessells & Monteiro, 2006).

The results of this study shed additional empirical light on a debate that has, to date, been grounded largely in clinical and field experience as well as in competing ideologies. Our findings suggest some clear, if somewhat nuanced conclusions. First, although our data do not speak to the clinical utility or cultural salience of PTSD as a diagnostic construct in Sri Lanka, the syndrome was clearly present in our sample and was significantly related to disaster exposure as well as to daily stressors. Currently, we do not know whether PTSD represents the most meaningful expression of posttraumatic distress among Sri Lankan youth. However, PTSD was sensitive to the predictors of interest in this study. Second, although war and disaster did exert a direct effect on psychological and psychosocial health among the youth in our sample, material deprivation, physical and sexual abuse, and witnessing interparental conflict also represented major sources of ongoing stress. With specific regard to PTSD, daily stressors accounted for significantly greater variance in PTSD levels, indicating that there are critical sources of traumatic stress affecting youth in eastern Sri Lanka other than their history of exposure to the conflict and the tsunami.

Unlike prior exposure to the conflict or the tsunami, daily stressors represent *current* and proximal sources of stress in children's lives and may be targeted for change. Our data strongly suggest several potentially effective strategies for improving young people's psychosocial well-being in this and other regions of Sri Lanka, including the development or expansion of child abuse prevention programs and greater attention to meeting the basic material needs of children and youth. Having adequate access to water, safe housing, and schooling are all related to children's well-being. Focus group discussions with children in refugee camps adjacent to swamps indicated that one of the greatest sources of stress for these children was the presence of poisonous snakes, which restricted people from going for water after dark, and generated considerable fear when walking to the toilets at night. This sort of proximal threat loomed considerably larger for children than memories of the tsunami that had occurred several months earlier.

Limitations

There are several limitations in this study that suggest caution when interpreting the results. The first is the use of a composite score of war- and tsunami-related stress. It could be argued that war-related events lead to different types of psychological outcomes, compared to tsunami-related events, and that a composite score may not capture how exposure to a single disastrous event could result in specific distress for a child. Although a significant linear relation was obtained between the composite score and outcomes, future research could examine which particular types of events are associated with specific outcomes. In this study, we wished to point more to the general patterns of results. Another limitation is that by conducting the study in schools, we may have excluded children who were most affected by the war and the tsunami, such that they were unable to attend school. Teachers did indicate that all children attended school regularly, and attendance records demonstrated over 98% school attendance overall, but it is possible that the children absent on the day the survey was administered may have been systematically different from the children who completed the survey, biasing the data.

Caution should also be used in generalizing the findings to all Sri Lankan youth or to a specific ethnocultural group, as the data were gathered in a particular setting where both the tsunami and the war had occurred. Lower reliability on some subscales of the measures also suggests caution in overinterpreting results until more studies are conducted that establish the relations found here. However, lower Cronbach's alpha is likely to underestimate reliability for the event-based scales used in this study. Because exposure to various disaster experiences does not necessarily co-occur, Cronbach's measure of internal consistency may be less relevant for such scales. More importantly, as Netland (2005) argues, war (and by extension, natural disaster) exposure scales should include the full range of events to which people may be exposed in a particular region. Our study did include such a range. Future research could also use teacher and parent reports against which self-report data could be compared, providing a more appropriate index of reliability for such measures of exposure and overcoming the added limitation of reliance on self-report.

Because we relied on the ability of children to read and comprehend the survey, younger children were not included in this study. Future research could include a wider age range, using methods such as stories and pictures to obtain similar information from younger children. The lack of longitudinal data also limits findings. The current study does not speak to how the developmental trajectory of children who report more distress may change as a function of exposure to disaster-related stress, and interfere with developmental tasks of childhood and adolescence, creating ongoing stress for these youth.

Adding other ecological dimensions such as parent characteristics and other cultural variables may have provided a richer profile of these youth, although many of these dimensions were captured in the SLIPSS–C. Future research could also examine how socioeconomic conditions might function as both mediators and moderators of outcomes for youth living in emergency settings. In a study conducted with Sri Lankan businessmen, de Mel, McKenzie, and Woodruff (2008) found that time, rather than economic recovery, was a better predictor of psychological recovery after trauma. Whether this is true for Sri Lankan children as well remains to be examined. In relation to this issue, although our model suggests that exposure to disasters predicts exposure to daily stressors, it is possible that a third variable (such as poverty) might be related to both, as well as to outcomes. It is not uncommon to find that far more poor than wealthy people reside in places where disasters typically strike. Thus, poverty may, on the one hand, place a child at risk for being in an undesirable location and, on the other, exacerbate the ill effects of exposure to stressors. The current model is only one perspective on the relationship between stressors and distress.

Despite these limitations, this study provides empirical evidence for the importance of examining daily stressors as significant contributors to the psychological and psychosocial distress of youth living in complex emergency settings, using culturally grounded measures. Although Sri Lanka's armed conflict is now officially ended, it must not be forgotten that its impact continues to be felt by Sri Lankans, particularly those living in the north and the east (CIDA, 2009; United Press International, 2008). Children and families there are struggling simply to survive. The results obtained in this study argue for more holistic and ecologically sound intervention programs that target specific stressors among Sri Lankan youth, to effect change and ameliorate the distress of this very vulnerable population.

References

Al-Krenawi, A., Lev-Wiesel, R., & Sehwail, M. (2007). Psychological symptomatology among Palestinian children living with political violence. *Child and Adolescent Mental Health*, 12, 27–31.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. Text Revision). Washington, DC: Author.

Amnesty International. (2009). *Amnesty International Report, 2009: Sri Lanka*. Retrieved May 6, 2010, from http://thereport.amnesty.org/en/regions/asia-pacific/sri-lanka

Barenbaum, J., Ruchkin, V., & Schwab-Stone, M. (2004). The psychosocial aspects of children exposed to war: Practice and policy initiatives. *Journal of Child Psychology and Psychiatry*, 45, 41–62.

Belfer, M. (2006). Caring for children and adolescents in the aftermath of disasters. *International Review of Psychiatry, 18*, 523–528.

Birleson, P., Hudson, I., Grey-Buchanan, D., & Wolff, S. (1987). Clinical evaluation of a self-rating scale for depressive disorder in childhood (Depression Self-Rating Scale). *Journal of Child Psychology and Psychiatry, 28*, 43–60.

Bonanno, G. (2008). The human capacity to thrive in the face of potential trauma. *Pediatrics, 121*, 369–375.

Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental processes. In W. Damon & R. Lerner (Eds.), *Handbook of developmental psychology* (Vol. 1, pp. 993–1028). Hoboken, NJ: Wiley.

Buckner, J. C., Bassuk, E. L., & Beardslee, W. R. (2004). Exposure to violence and low-income children's mental health: Direct, moderated and mediated relations. *American Journal of Orthopsychiatry, 74*, 413–423.

Cairns, E., & Dawes, A. (1996). Ethnic and political violence: A commentary. *Child Development, 67*, 129–139.

Canadian International Development Agency. (2009). *Status report: Post-tsunami situation, Ampara district* Retrieved May 6, 2010, from http://www.humanitarianinfo.org/srilanka/infocentre/assessments/others/doc/CIDA/Status%20%20Verification%20Report%20on%20Ampara%20by%20Cida%20Canada.pdf

Cicchetti, D., & Lynch, M. (1993). Toward an ecological/transactional model of community violence and child maltreatment: Consequences for children's development. *Psychiatry, 56*, 96–118.

Cooper, S., & Guthrie, B. (2007). Ecological influences on health-promoting and health-compromising behaviors: A socially embedded approach to urban African American girls' health. *Family and Community Health, 30*, 29–41.

Davis, L., & Siegel, L. J. (2005). Posttraumatic stress disorder in children and adolescents: A review and analysis. *Clinical Child Family Psychology Review, 3*, 135–154.

Dawes, A., & Donald, D. (1994). Understanding the psychological consequences of adversity. In A. Dawes & D. Donald (Eds.), *Childhood and adversity* (pp. 1–27). Claremont, South Africa: David Philip.

de Berry, J., Fazili, A., Said, F., Nasiry, F., Hashemi, S., & Hakimi, M. (2003). *The children of Kabul: Discussions with Afghan families*. Kabul, Afghanistan: Save the Children USA and UNICEF.

de Mel, S., McKenzie, D., & Woodruff, C. (2008). Mental health recovery and economic recovery after the tsunami: High-frequency longitudinal evidence from Sri Lankan small business owners. *Social Science and Medicine, 3*, 582–595.

Department of Census and Statistics, Sri Lanka. (2001). *Number and percentage of population by district and ethnic group*. Retrieved June 3, 2008, from http://www.statistics.gov.lk

Department of Census and Statistics, Sri Lanka. (2005). *Impact of tsunami 2004 on Sri Lanka*. Retrieved January 4, 2008, from http://www.statistics.gov.lk/Tsunami/

Durvasula, R., & Mylvaganam, G. (1994). Mental health of Asian Indians: Relevant issues and community implications. *Journal of Community Psychology, 22*(2), 97–108.

Evans, S., Davies, C., & DiLillo, D. (2008). Exposure to domestic violence: A meta-analysis of child and adolescent outcomes. *Aggression and Violent Behavior, 13*, 131–140.

Fazel, M., Wheeler, J., & Danesh, J. (2005). Prevalence of serious mental disorders in 7,000 refugees resettled in Western countries: A systematic review. *Lancet, 365*, 1309–1314.

Fernando, G. A. (2005). Interventions for survivors of the tsunami disaster: Report from Sri Lanka. *Journal of Traumatic Stress, 17*, 267–268.

Fernando, G. A. (2008). Assessing mental health and psychosocial status in communities exposed to traumatic events: Sri Lanka as an example. *American Journal of Orthopsychiatry, 78*, 229–239.

Fernando, G., & Makanui, P. K. (2005, June). Children exposed to violence: Mental health and self-esteem. In K. Putnam (Chair), *Trauma exposure and its impact on children*. Symposium conducted at the annual meeting of the European Society for Traumatic Stress Studies, Stockholm, Sweden.

Fernando, G. A., Miller, K. E., Uyanavithanage, S., Prasad, J., & Garcia, J. (2008, August). *The impact of exposure to traumatic events among children in Sri Lanka*. Poster presented at the annual meeting of the American Psychological Association, Boston.

Fitzgerald, H. E., Karraker, K. H., & Luster, T. (Eds.). (2002). *Infant development: Ecological perspectives*. New York: Routledge Falmer.

Foa, E. B., Johnson, K., Feeny, N., & Treadwell, K. R. (2001). The Child PTSD Symptoms Scale: A preliminary examination of its psychometric properties. *Journal of Clinical Child Psychology, 30*, 376–384.

Garbarino, J., & Kostelny, K. (1996). The effects of political violence on Palestinian children's behavior problems: A risk accumulation model. *Child Development, 67*, 33–45.

Hart, J., Galappatti, A., Boyden, J., & Armstrong, M. (2007). Participatory tools for evaluating psychosocial work with children in areas of armed conflict: A pilot in eastern Sri Lanka. *Intervention, 5*, 41–60.

Hawkins, S. S., & Radcliffe, J. (2006). Current measures of PTSD for children and adolescents. *Journal of Pediatric Psychology, 31*, 420–430.

Heptinstall, E., Sethna, V., & Taylor, E. (2004). PTSD and depression in refugee children: Association with pre-migration trauma and post-migration stress. *European Journal of Child and Adolescent Psychiatry, 13*, 373–380.

Hollifield, M., Warner, T. D., Lian, N., Krakow, B., Jenkins, J., Kesler, J., et al. (2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association, 288*, 611–621.

Inter-Agency Standing Committee. (2007). *IASC guidelines on mental health and psychosocial support in emergency* *settings*. Retrieved June 3, 2008, from http://www.humanitarianinfo.org/iasc/pageloader.aspx?page=content-products-products&productcatid=22

International Crisis Group. (2007, May). *Sri Lanka's Muslims: Caught in the crossfire (Asia Report No. 134)*. Colombo, Sri Lanka/Brussels, Belgium: International Crisis Group.

Kalksma-Van Lith, B. (2007). Psychosocial interventions for children in war-affected areas: The state of the art. *Intervention, 5*, 3–17.

Kilpatrick, D. (2004). The ethics of disaster research. *Journal of Traumatic Stress, 17*, 361–362.

Macksoud, M., & Aber, L. (1996). The war experiences and psychosocial development of children in Lebanon. *Child Development, 67*, 70–88.

March, J. S., Parker, J., Sullivan, K., Stallings, P., & Conners, K. C. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 554–565.

Masten, A. S. (2006). Developmental psychopathology: Pathways to the future. *International Journal of Behavioral Development, 30*, 47–54.

Miller, K. E., & Fernando, G. A. (2008). Epidemiological assessment in emergency settings: Recommendations for enhancing a potentially useful tool. *Intervention, 6*, 255–260.

Miller, K. E., Fernando, G. A., & Berger, D. E. (2009). Daily stressors in the lives of Sri Lankan youth: A mixed methods approach to assessment in a context of war and natural disaster. *Intervention, 7*, 187–203.

Miller, K. E., Fernando, G. A., Uyanavithanage, S., Prasad, J., & Ranawake, M. A. J. (2008). *The relative contribution of daily stressors and disasters to the psychosocial status of Sri Lankan youth*. Paper presented at the International Conference on Community Psychology, Lisbon, Portugal.

Miller, K. E., Kulkarni, M., & Kushner, H. (2006). Beyond trauma-focused psychiatric epidemiology: Bridging research and practice with war-affected populations. *American Journal of Orthopsychiatry, 76*, 409–422.

Miller, K. E.., Omidian, P., Kulkarni, M., Yaqubi, A., Daudzi, H., & Rasmussen, A. (2009). The validity and clinical utility of post-traumatic stress disorder in Afghanistan. *Transcultural Psychiatry, 46*, 219–297.

Miller, K. E., Omidian, P., Rasmussen, A., Yaqubi, A., & Daudzai, H. (2008). Daily stressors, war experiences, and mental health in Afghanistan. *Transcultural Psychiatry, 45*, 611–638.

Miller, K. E., & Rasco, L. M. (2004). An ecological framework for addressing the mental health needs of refugee communities. In K. E. Miller & L. M. Rasco (Eds.), *The mental health of refugees: Ecological approaches to healing and adaptation* (pp. 1–64). Mahwah, NJ: Erlbaum.

Montgomery, E. (2008). Self- and parent assessment of mental health: Disagreement on externalizing and internalizing behaviour in young refugees from the Middle East. *Clinical Child Psychology and Psychiatry, 13*, 49–63.

Myers, K. M., & Winters, N. C. (2002). Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 634–659.

Netland, M. (2005). Event-list construction and treatment of exposure data in research on political violence. *Journal of Traumatic Stress, 18*, 507–517.

Neuner, F., Schauer, E., Catani, C., Ruf, M., & Elbert, T. (2006). Post-tsunami stress: A study of post-traumatic stress disorder in children living in three severely affected regions in Sri Lanka. *Journal of Traumatic Stress, 19*, 339–347.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., Kaniasty, K., et al. (2002). 60,000 disaster victims speak: Part I. *Psychiatry, 65*, 207–239.

Osofsky, J. D. (Ed.). (2004). *Young children and trauma*. New York: Guilford.

Oyserman, D., Coon, H., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72.

Pynoos, R. S., Steinberg, A. M., & Piacentini, J. C. (1999). A developmental psychopathology model of childhood traumatic stress. *Biological Psychiatry, 46*, 1542–1554.

Qouta, S., Punamäki, R., & El Sarraj, E. (2003). Prevalence and determinants of PTSD among Palestinian children exposed to military violence. *European Child and Adolescent Psychiatry, 12*, 265–272.

Rasmussen, A., Nguyen, L., Wilkinson, J., Raghavan, S., Vundla, S., Miller, K., et al. (2010). Rates and impact of trauma and current stressors among Darfuri refugees in Eastern Chad. *American Journal of Orthopsychiatry, 80*, 223–232.

Shaw, J. A. (2000). Children, adolescents and trauma. *Psychiatry Quarterly, 71*, 227–243.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. *Sociological Methodology, 13*, 290–312.

Somasundaram, D. (2003). Collective trauma in Sri Lanka. *Intervention: International Journal of Mental Health, Psychosocial Work and Counselling in Areas of Armed Conflict, 1*, 4–13.

South Asia Free Media Association. (2008). *Sri Lanka: Ethnic conflict, civil war, and politics*. Retrieved June 3, 2008, from http://www.lankalibrary.com/pol/militancy_history.htm

Stichick, T. (2001). The psychosocial impact of armed conflict on children. Rethinking traditional paradigms in research and intervention. *Child and Adolescent Psychiatric Clinics of North America, 10*, 797–814.

Summerfield, D. (1999). A critique of seven assumptions behind psychological trauma programmes in war-affected countries. *Social Science and Medicine, 48*, 1449–1462.

Thabet, A., & Vostanis, P. (2000). Post-traumatic stress disorder reactions in children of war: A longitudinal study. *Child Abuse and Neglect, 24*, 291–298.

Triandis, H. C. (Series Ed.). (1995). *New directions in social psychology: Individualism and collectivism.* Boulder, CO: Westview Press.

UNICEF. (2008). *State of the world's children.* Retrieved June 3, 2008, from http://www.unicef.org/sowc/index_sowc.html

United Press International. (2008). *All is not normal in Eastern Sri Lanka.* Retrieved January 11, 2009, from http://www.upiasia.com/Politics/2008/12/17/all_is_not_normal_in_eastern_sri_lanka/4151/

United States Committee for Refugees. (2007). *World Refugee Survey: 2007.* Retrieved December 12, 2009, from http://www.refugees.org/article.aspx?id=2324&subm=179&area=About%20Refugees&

Wessells, M., & Monteiro, C. (2006). Psychosocial assistance for youth: Toward reconstruction for peace in Angola. *Journal of Social Issues*, *62*, 121–139.

Yule, W. (2002). Alleviating the effects of war and displacement on children. *Traumatology*, *8*, 160–180.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1211–1227

Exploring Posttraumatic Growth in Children Impacted by Hurricane Katrina: Correlates of the Phenomenon and Developmental Considerations

Ryan P. Kilmer and Virginia Gil-Rivas
University of North Carolina at Charlotte

This study explored posttraumatic growth (PTG), positive change resulting from struggling with trauma, among 7- to 10-year-olds impacted by Hurricane Katrina. Analyses focused on child self-system functioning and cognitive processes, and the caregiving context, in predicting PTG at 2 time points (Time 1 n = 66, Time 2 n = 51). Findings suggest that rumination, both negative, distressing thoughts and constructive, repetitive thinking, plays an important role in PTG. Hypotheses regarding future expectations and perceived competence were not fully supported, and, unexpectedly, coping competency beliefs, realistic control attributions, and perceived caregiver warmth did not contribute to PTG models. With 1 exception (positive reframing coping advice), caregiver-reported variables did not relate to PTG; no caregiver variable reached significance in final models. Relevant theory, developmental considerations, and future directions are discussed.

The child disaster literature has largely focused on symptomatology and problem behaviors, primarily posttraumatic stress symptoms (PTSS) and posttraumatic stress disorder (PTSD), and factors that appear to influence children's adjustment (Silverman & La Greca, 2002). While that work has yielded necessary findings about the consequences of disaster and informed subsequent research, intervention, and policy, this evolving field needs to go beyond negative sequelae and elucidate factors that may mitigate the impact of disaster, as well as factors and conditions that contribute to resilience and processes such as *posttraumatic growth* (PTG; Alisic, van der Schoot, van Ginkel, & Kleber, 2008; Kilmer & Gil-Rivas, 2010).

PTG refers to positive change experienced as a result of the struggle with trauma. Emphasizing the transformative elements of responding to adversity, PTG has been the focus of burgeoning interest in the research and clinical literatures (e.g., Calhoun & Tedeschi, 2006; Helgeson, Reynolds, & Tomich, 2006). However, despite this attention, and the significant research base regarding children and youth's response to stress or trauma (Compas, Connor-Smith, Saltzman, Harding Thomsen, & Wadsworth, 2001; Grant et al., 2003; Osofsky, 2004a), relatively few studies have examined the construct among children. Notably, the report of the Hurricane Katrina Community Advisory Group (HKCAG, 2006) assessed PTG; however, it did so only among adults.

Nevertheless, early evidence supports the conclusion that growth occurs in some form among children and youth who have experienced natural disasters (Cryder, Kilmer, Tedeschi, & Calhoun, 2006; Kilmer et al., 2009), terrorism (e.g., Laufer & Solomon, 2006), traffic accidents (Salter & Stallard, 2004), cancer (e.g., Barakat, Alderfer, & Kazak, 2006), and a range of traumatic events (e.g., Alisic et al., 2008; Ickovics et al., 2006; Milam, Ritt-Olson, & Unger, 2004). However, the literature appears to include only four published papers (Alisic et al., 2008; Cryder et al., 2006; Phipps, Long, & Ogden, 2007; Salter & Stallard, 2004) investigating PTG or related constructs (stress-related growth, perceived benefits, benefit finding) within a child sample or subsample. Of those, only Cryder et al. (2006) reported correlates of PTG reflecting resources, competencies, or processes thought to relate to the growth process. This article seeks to extend those findings and enhance understanding of PTG in children. It has the following objectives: (a) provide relevant conceptual background, considering developmental issues, distinguishing the construct from resilience, and discussing selected variables thought to be relevant to the PTG process and (b) examine

---

This research was supported by National Institute of Mental Health Award R03 MH078197-01 (V.G.-R. and R.P.K.).

Correspondence concerning this article should be addressed to Ryan P. Kilmer, Department of Psychology, The University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223-0001. Electronic mail may be sent to rpkilmer@uncc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0013

correlates of PTG. Specifically, informed by Kilmer's (2006) model of the growth process and ideas posited by Kilmer and Gil-Rivas (2010), the present work focuses on the role of child characteristics, including self-system variables (e.g., perceived competence, future expectations) and cognitive resources and processes (e.g., realistic control expectations, rumination), and indicators of the caregiving context (e.g., caregiver warmth, caregiver positive reframing coping advice) thought, on theoretical and/or empirical grounds, to be related to PTG.

PTG and the Growth Process

Although many children adapt successfully in the face of adversity, research has demonstrated that children who have experienced trauma may exhibit difficulties in multiple developmental domains, such as emotional and behavioral symptoms, delays or regressions in developmental skills, and difficulties in cognitive and socioemotional functioning (e.g., Cicchetti & Lynch, 1993). Disaster-specific work has reinforced such notions, suggesting that a sizable proportion of children evidence a range of short- and long-term mental health sequelae following disaster (e.g., La Greca, Silverman, Vernberg, & Roberts, 2002; Pfefferbaum et al., 1999), which vary based on a child's developmental level (see National Child Traumatic Stress Network [NCTSN], n.d.; Osofsky, Osofsky, & Harris, 2007). The most common long-term negative consequences reflect PTSD symptoms, anxiety and generalized distress, major depression, suicidality, and problem behaviors (e.g., Silverman & La Greca, 2002).

The child trauma literature has also described findings akin to PTG. For instance, work with pediatric cancer survivors suggested that youth experienced positive changes including heightened sensitivity to others, increased altruism, and modified values (Fritz, Williams, & Amylon, 1988). A recent study with youth aged 7–18 years who had experienced traffic accidents found that many reported a changed philosophy of life, improved relationships, and more positive self-perceptions after the accidents (Salter & Stallard, 2004). Others have documented instances of children reporting that they feel ''stronger'' after significant life events, such as the death of a parent (Krementz, 1983), and more systematic efforts, focusing on PTG, have found evidence supporting the phenomenon in children (Alisic et al., 2008; Cryder et al., 2006; Kilmer et al., 2009). In general, the literature suggests that the positive changes framed as PTG tend to cohere in several main domains: a greater sense of one's personal strength; a different perspective on one's relationships; a changed philosophy of life, such as a greater appreciation for life and its new possibilities; and spiritual growth (Calhoun & Tedeschi, 2006; Kilmer, 2006).

Key elements of the hypothesized PTG process warrant mention (see Kilmer, 2006). In many cases, traumas such as disaster can impact meaningfully a youngster's assumptive world. They can shatter assumptions about one's self, others, one's world, and the expected course of one's life (Janoff-Bulman, 1992), stimulating attempts to cope and adapt. Growth is thought to evolve as a result of the struggle with trauma and its aftermath, not solely the experience of the trauma itself, developing as one grapples with his or her new reality, working to understand what has happened and its implications for life going forward (Calhoun & Tedeschi, 2006). In fact, theorists postulate that this continuing distress and the individual's efforts to reconcile his or her posttrauma reality facilitate a constructive cognitive processing of trauma, or productive rumination (Calhoun & Tedeschi, 2006). Through this process, one may try to make sense of the event(s) and integrate the trauma and its aftermath in a manner consistent with prior internal representations, or working models. Subsequently, this ruminative process yields schema change, a critical element that serves to consolidate changed perspectives on self, others, and one's new life and way of living (Calhoun & Tedeschi, 2006). Of relevance here, existing evidence suggests that, following trauma, the reactions and responses of children and youth at different ages and stages vary, in part because their cognitive and emotional capacities lead them to understand and internalize the experience differently (NCTSN, n.d.; Osofsky, 2004b). In light of these realities, and because PTG appears to involve a degree of cognitive sophistication that allows both losses and gains to be recognized, the degree to which the growth process in children accords with the process observed among adults is unclear (Cryder et al., 2006).

Developmental Considerations in the Study of PTG

Questions have been raised about children's capacity to engage in the cognitive and affective processes necessary to yield transformational changes, including their ''psychological mindedness'' (intrapersonal awareness and insight) or self-understanding (Cryder et al., 2006; Kilmer, 2006). Variability in youngsters' cognitive capabilities may influence their encoding and appraisal of, and attributions about, trauma (Hasan & Power, 2004; Shahinfar & Fox, 1997), repertoire of coping strategies, ability to marshal resources and cope effectively, and capacity to attend to and report their internal experiences and emotional states (Alisic et al., 2008; Cryder et al., 2006; Kilmer, 2006; Osofsky, 2004b). Moreover, children's responses and reactions are closely tied to their cognitive, emotional, and behavioral capacities at the time of the trauma and the time of the expression of their responses posttrauma (Shahinfar & Fox, 1997). Although this study does not address these issues directly, several developmental considerations warrant a brief discussion, particularly the degree to which child self-reported PTG reflects actual growth versus normative maturation, and cognitive maturity and the appraisal and understanding of the event, including issues of self-awareness and self-understanding.

Some researchers have underscored the need to differentiate child PTG from change resulting from normal maturational processes (Cohen, Hettler, & Pane, 1998). Two recent efforts shed light on this valid issue. Taku, Calhoun, Kilmer, and Tedeschi (2008) found that, among Japanese youth (mean age = 13.52 years, *SD* = 0.97), those who reported experiencing at least one trauma or major stressful event within the past year scored significantly higher on a measure of PTG (and on 8 of its 10 items) than peers who had not (who, in response to the PTG measure's items, reported about the degree to which they had changed in the last year). In their community sample (mean age = 10.24 years, *SD* = 1.21), Alisic et al. (2008) asked children to describe their "worst experience ever" and to consider that experience as they answered items about PTSS and PTG. An expert panel then determined whether the events reported were consistent with the *DSM–IV* PTSD A1 criterion for objective trauma exposure (i.e., "the person experienced, witnessed, or was confronted with an event or events that involve actual or threatened death or serious injury; or threat to the physical integrity of himself or herself or others"; American Psychiatric Association [APA], 1994, p. 427) and classified children on the basis of their experiences (i.e., exposed or not exposed to trauma). Consistent with Taku et al., children exposed to trauma reported more PTG than those without trauma experiences (Alisic et al., 2008). While prospective-longitudinal methodologies are necessary to yield compelling evidence that the growth assessed is indeed PTG, these findings provide support for the PTG process as distinct from normal growth.

When considering cognitive maturity, some may question whether young children are sufficiently self-aware to reflect and consider the degree to which they have changed over time. Kilmer et al. (2009) reported responses to findings from open-ended items illustrating that, while few 7- to 10-year-olds spontaneously made comments that clearly suggested growth (e.g., "I am more positive now"), they demonstrated a grasp of the temporal element of the questions about change (e.g., "After the hurricanes, my grandfather died and my mother got meaner"). This capacity to make the temporal comparison is consistent with prior findings (Harter, 2006).

Beyond the temporal element, with cognitive maturity comes the capacity for greater self-understanding and awareness and more developed schema for oneself and one's world, key factors in the processes presumed relevant to growth. Findings regarding the impact of age have been mixed: Milam et al. (2004) found a positive association between age and PTG among adolescents, positing that a particular level of cognitive maturity is necessary to find meaning or identify salient positive changes as a result of trauma; however, in much younger samples, Cryder et al. (2006) and Kilmer et al. (2009) did not detect a relation between age and PTG.

Nevertheless, multiple salient capacities emerge in middle and late childhood that warrant mention, given their potential relevance to PTG. For example, during this time, children evidence increased competence in managing trauma-related thoughts and regulating their emotions (e.g., Salmon & Bryant, 2002). Thus, authors have noted an increase in emotion-focused coping and the manifestation of cognitively oriented attempts at coping, including cognitive reframing, or reappraising a difficult situation (e.g., Aldwin, 2007; Compas et al., 2001; Salmon & Bryant, 2002). Moreover, although findings vary regarding the ages at which this capacity coalesces, between ages 6 and 9, control expectations become more realistic, and children are better able to identify the ways in which they can cope with those circumstances that can and cannot be controlled (e.g., Aldwin, 2007; Salmon & Bryant, 2002).

Furthermore, given that (a) schema change and (b) the capacity to acknowledge and integrate the positives in oneself and one's situation following trauma constitute two critical elements of the hypothesized PTG process, it is necessary to highlight that young children's internal representations of self and others are not yet set and may be qualitatively different than those of adults. According to

Janoff-Bulman (1992), because a child's assumptive world is more pliable than an adult's, his or her basic assumptions and working models may be modified as new experiences become incorporated. Thus, exposure to trauma can severely disrupt one's assumptive world (e.g., of safety, predictability), because these cumulatively constructed representations are less firmly entrenched (Janoff-Bulman, 1992; Lieberman & Van Horn, 2004). However, that also suggests that the child's schema may be open to adaptive inputs (Janoff-Bulman, 1992), including the reworking and revision of one's views of oneself and one's world.

For that to occur, the child must recognize and internalize some positive appraisals of the trauma and subsequent changes. A line of research regarding children's ability to acknowledge the simultaneous existence of positive and negative emotions or attributes in relation to a target event (Harter, 1986; Harter & Buddin, 1987) may shed light on their capacities to accommodate a positive aspect or change in the face of trauma. Specifically, while 5- to 7-year-olds are generally unable to integrate emotions of opposing valence (both positive and negative), a developmental progression is evident such that children develop the capacity to understand and express emotions of different valence to a target between 9 and 13 years of age ($M = 11.34$, $SD = 2.12$), with some children evidencing this ability as young as 7.6 years (Harter, 1986; Harter & Buddin, 1987). These findings may suggest a lower age limit for PTG (Kilmer, 2006).

The PTG and Resilience Constructs: Parallel but Distinct

PTG clearly shares significant conceptual variance with the construct of resilience; however, they are distinct. While resilience refers to "a dynamic developmental process reflecting evidence of positive adaptation despite significant life adversity" (Cicchetti, 2003, p. xx; see Masten, 2001), PTG refers to a process by which one is *transformed* as a result of his or her struggle in trauma's aftermath (Cryder et al., 2006). Thus, PTG involves positive changes extending beyond sound adjustment despite adversity; it "involves movement beyond pretrauma adaptation, a qualitative change in functioning across domains" (Cryder et al., 2006, pp. 65–66).

That PTG does not equate to positive adjustment further distinguishes PTG and resilience. That is, those reporting PTG may also report less emotional well-being or positive adjustment than those evidencing resilient adaptation (Cryder et al., 2006). Although a recent prospective-longitudinal study with adolescents demonstrated that those who reported higher PTG exhibited less emotional distress up to 12 and 18 months postevent (Ickovics et al., 2006), and others have reported a link between PTG and health-related behaviors (Milam et al., 2004), the extant literature largely suggests that while many survivors recognize positive change as a result of their struggle, they often report distress and symptomatology in the aftermath of trauma (Calhoun & Tedeschi, 2006). Thus, growth and distress may coexist (Kilmer & Gil-Rivas, 2008). In that vein, Linley and Joseph (2004) suggested framing growth and distress as independent dimensions, not two ends of a single continuum. Accordingly, Salter and Stallard (2004) reported that while 42% of their sample reported PTG, 37% of these same children met criteria for a PTSD diagnosis. Similarly, Alisic et al. (2008) and Kilmer et al. (2009) found a significant positive correlation ($rs = .41$ and $.44$, respectively) between PTG and PTSS. Such findings are not inconsistent with PTG models—distress is viewed as a key catalytic element of the process and may serve to maintain growth (Tedeschi, Calhoun, & Cann, 2007).

Notwithstanding the differences between the constructs, the resilience research base can point to factors, processes, and conditions that may not only influence children's adjustment in response to trauma, including disaster, but may also hold promise in the study of PTG. For instance, most disaster studies have focused on relatively circumscribed individual characteristics (e.g., demographics) and elements of functioning, such as coping approaches (e.g., La Greca, Silverman, Vernberg, & Prinstein, 1996) or predisaster functioning (e.g., anxiety symptoms; see, e.g., Weems et al., 2007). Although these studies have yielded important findings, work in resilience suggests a far wider range of child variables that warrants investigation, particularly child attributes, resources, and competencies such as positive dispositional qualities, good intellectual functioning, self-efficacy, perceived competence, control attributions, and positive future expectations or optimism (e.g., Luthar, 2003; Masten & Coatsworth, 1998). Results from this work may inform PTG research.

Constructs of Relevance to the Present Study

Drawing from prior PTG research and theory (e.g., Cryder et al., 2006; Kilmer, 2006), as well as work in stress and coping and risk and resilience (e.g., Luthar, 2003; Masten & Coatsworth, 1998), the following constructs are of prime relevance to the present study.

Self-System Factors: Perceptions Regarding Competence and the Future

One's perceptions of competence, one's expectations for his or her future, and one's positive appraisals and expectations regarding one's ability to meet daily task demands, cope, and adjust in the face of stress or trauma are hypothesized to play a core role in the PTG process (Kilmer, 2006). Cryder et al. (2006) found a direct relation between competency beliefs and PTG and, in the resilience literature, positive self-views and self-system functioning are among the most consistently reported factors that seemingly serve a protective function, facilitating positive adaptation and reducing risk of maladjustment in the face of adversity (e.g., Luthar, Cicchetti, & Becker, 2000; Masten, 2001). For instance, perceived competence scores were among the most powerful discriminators of those demonstrating resilient versus maladjusted outcomes in two independent cohorts of children exposed to multiple risks (Cowen et al., 1992; Hoyt-Meyers et al., 1995). Additionally, in their influential work, Werner and Smith (1982) emphasized that those exhibiting resilient adaptation ''worked well, played well, loved well, and *expected* well'' (p. 153, italics added). Positive beliefs about one's competencies or future may influence how a youngster perceives an event, his or her responses to adversity, and the effort he or she sustains in wrestling with the event and its aftermath (e.g., Wyman, Cowen, Work, & Kerley, 1993). In turn, factors reflecting sound self-system functioning (e.g., future expectations, perceived competence) are hypothesized to relate directly to PTG because they may affect not only one's perceptions of these events and ability to deal with them but also one's approach to and persistence in coping and the likelihood of perceiving positive change (Kilmer, 2006).

Cognitive Resources: Realistic Control Expectations

Current PTG models highlight cognitive processes, including not only appraisals of the traumatic event but also control attributions and rumination (Calhoun & Tedeschi, 2006; Kilmer, 2006). The accuracy of children's control expectations can influence how they respond and react following adversity (see Wannon, 1990). Wyman (2003) argued that under adversity, realistic control—that is, accurate, age-appropriate perceptions about events, outcomes, and problems that are controllable and uncontrollable—may be more meaningful and adaptive for children than an undifferentiated internal locus of control, a hypothesis with some support (Cowen et al., 1992; Hoyt-Meyers et al., 1995). For example, children's realistic control expectations and attributions about their circumstances would seem salient in the postdisaster context, as they may influence their responses to adversity, coping approaches, affect regulation, and behavioral organization and goals (Kilmer, 2006; see Hasan & Power, 2004; Pynoos, Steinberg, & Wraith, 1995). In turn, Kilmer (2006) hypothesized that realistic control relates bidirectionally with other cognitive elements of the growth process, such as productive rumination, and with aspects of the self-system, including coping competency beliefs and hope for the future.

Productive Rumination

The function of rumination, or repetitive thought, is thought to be of particular import in the PTG process. Ruminative thinking can have negative consequences (intrusive rumination has been linked with depressive symptoms and anxiety) or can serve a more constructive function, for example, helping one to recover and perhaps even grow posttrauma (see Watkins, 2008). According to PTG models, disruptions to one's assumptive world and distress associated with trauma (including intrusive thinking) may result in productive, or deliberate, ruminative processes that are thought to help one better understand, engage, and perhaps find meaning in the event, allowing for the inclusion of positive appraisals. These positive appraisals facilitate schema change as one notes and integrates the changes in oneself and one's environment (Kilmer, 2006; Tedeschi et al., 2007). Kilmer (2006) hypothesized a link between deliberate ruminative thinking and self-system variables, and consistent with this notion, Cryder et al. (2006) reported a significant relation between rumination and children's competency beliefs, with the latter the prime proximal predictor of PTG. While the nature of the rumination–PTG relation is not clear, early indications suggest the amount of rumination alone is not a significant predictor of PTG (Cryder et al., 2006). It is possible the content of the repeated thoughts and the manner in which the individual engages them more actively contribute to, and perhaps catalyze, the growth process (Kilmer, 2006; Tedeschi et al., 2007).

The Caregiving Context

Caregivers and, more broadly, the caregiving context are widely recognized to play a central role in helping children adapt in the face of adversity (Luthar, 2003; Masten & Coatsworth, 1998) and, more specifically, disasters (Gil-Rivas, Holman, & Silver, 2004). Among many potentially salient elements of this context, caregiver functioning postdisaster (including distress and symptoms) may relate to their responsiveness and availability (affecting caregiver behaviors and support) as well as children's appraisals, interpretations, reactions, and behavior (Masten, Best, & Garmezy, 1990; Salmon & Bryant, 2002).

Caregivers can also help youngsters appraise and understand events and their emotions, provide support, share perspective, and guide or model coping (Gil-Rivas, Silver, Holman, McIntosh, & Poulin, 2007; Kliewer, Sandler, & Wolchik, 1994; Pynoos et al., 1995). In turn, coping advice involving positive reframing, that is, encouraging children to consider the positive, would seem germane to child PTG. In the same vein, caregiver PTG may evidence a meaningful link to child PTG, in part because their experiences may inform the manner by which they help the child craft narratives about, or increase understanding of, the changes that have occurred, and offer perspectives that can foster adaptive schema change (Calhoun & Tedeschi, 2006; Janoff-Bulman, 1992; Kilmer, 2006).

Researchers have emphasized that warm, nurturant caregiving can mitigate the impact of risk (Luthar et al., 2000; Masten & Coatsworth, 1998), and competent, developmentally appropriate parenting in the disaster's aftermath will facilitate attunement and responsiveness to children's needs (Kilmer, 2006). In addition, supportive and nurturant caregiver–child relationships are thought to positively influence children's self-views and self-system functioning (Yates, Egeland, & Sroufe, 2003). In the PTG literature, this possibility was borne out by the significant link identified between perceived support and children's coping competency beliefs (Cryder et al., 2006). Thus, within the PTG process, warm, responsive caregiving may contribute to PTG largely through its association with the self-system.

The Context of the Present Study: Children and Families Impacted by Hurricane Katrina

The adversities experienced and the consequences evidenced by the children and families impacted by Katrina have been well documented (Gil-Rivas, Kilmer, Hypes, & Roof, 2010; HKCAG, 2006; Osofsky, Osofsky, Kronenberg, & Cross, 2010; Weems et al., 2007). In describing the experiences of the sample studied in the present investigation, Gil-Rivas et al. (2010) noted that 86.8% endorsed experiencing at least one hurricane-related traumatic event, and 63.2% of children reported that Hurricane Katrina was one of the ''most scary'' events they had experienced. Overall, children reported experiencing considerable adversity and, in the face of such circumstances, 1-year post-Hurricane Katrina, 21% of the children obtained PTSS scores suggesting the presence of a probable PTSD diagnosis, and nearly all reported experiencing at least one PTSS much or most of the time during the previous month (Gil-Rivas et al., 2010).

Kilmer et al. (2009) examined PTG among those same children, validating the PTG measure and reporting that, at baseline (approximately 12 months postdisaster), roughly 50% of the sample obtained total scores, suggesting an average response of at least ''some'' perceived change, or growth, across the scale's items; 7.5% had scores suggesting average responses of ''a little'' change or less. Total PTG scores did not differ for boys and girls, and age did not correlate with PTG. Analyses examined the linkages between PTG and multiple trauma-related variables. Consistent with expectations, children's baseline PTG scores were not significantly associated with objective hurricane exposure, but children's self-reported subjective response to the event and self-reported PTSS correlated significantly with PTG. Regression analysis demonstrated that subjective response related to PTG over and above objective hurricane exposure, though, in the final model, only PTSS significantly explained variance in PTG scores. Similar findings held at follow-up, roughly 22 months posthurricane. For instance, 53% of the sample obtained PTG scores suggesting ''some'' perceived change or more; 13.2% had scores suggesting ''a little'' or less. Consistent with baseline results, there were no gender differences in PTG scores, and age did not relate to PTG. Although baseline PTSS correlated with PTG at follow-up, baseline PTG was the only significant predictor of PTG. Overall, these findings lay groundwork for the present study.

The Present Study

This study sought to extend previous PTG work (i.e., Cryder et al., 2006; Kilmer et al., 2009), enhance understanding of PTG, and expand its nomological network by exploring relations beyond those between PTG and trauma-related variables (e.g., subjective response, PTSS). Specifically, it examined associations between PTG, children's self-system and cognitive resources and processes, and the caregiving context in a sample of young children (aged 7–10 years) impacted by Hurricane Katrina. This age group was targeted because it would permit examination of relevant developmental considerations and change over time, particularly among key constructs hypothesized to be associated with PTG. Informed by prior research and theory (Cryder et al., 2006; Kilmer, 2006; Kilmer & Gil-Rivas, 2010), the age of the children (and their still-evolving self-systems), the presumed importance of the caregiving context, and the assumed role of ruminative thinking in catalyzing the PTG process, we expected that:

1. At baseline, self-system variables (perceived competence, future expectations, and coping competency beliefs) would relate positively to PTG.
2. At baseline, perceived caregiver warmth would relate to both self-system variables and PTG, caregiver distress would be negatively associated with PTG, and caregiver PTG and positive reframing coping advice would be positively associated with child PTG.
3. At baseline, cognitive variables, particularly deliberate rumination, would evidence a positive association with PTG, over and above caregiver and self-system variables.
4. Baseline levels of child self-system functioning, rumination (intrusive and deliberate), and caregiver's positive reframing coping advice would predict PTG at follow-up, over and above PTG at baseline.

Method

Participant Recruitment and Study Design

Face-to-face interviews were conducted with children and their primary caregiver in the Baton Rouge and New Orleans, Louisiana metropolitan areas and along the Mississippi Gulf Coast. Baseline interviews were conducted roughly 1 year post-Hurricane Katrina ($M = 12.62$ months, $SD = 4.08$); follow-up interviews were conducted approximately 2 years posthurricane ($M = 22.36$ months, $SD = 3.25$). Participants were recruited via flyers distributed at elementary schools, Federal Emergency Management Agency (FEMA)-operated trailer parks, service provider agencies, and community-based organizations, and via participant referral. Trained research assistants conducted the interviews at participants' homes or at locations convenient to caregivers. Participants were compensated with gift cards for their participation and that of their child. The study was approved by the University of North Carolina at Charlotte and the Louisiana State University Institutional Review Boards.

Participants

At baseline (T1), participants included 66 caregiver–child dyads; of these, 51 completed the follow-up (T2) interview (77% retention). At T1, children, on average, were 8.5 years of age ($SD = 1.1$), and 56.1% were female. A majority of children were African American (77.3%), 15.2% were White, and 7.5% reflected other backgrounds. Caregivers were predominantly female (87.9%), and on average, they were 38.1 years old ($SD = 9.6$); 80.9% of them were the biological parent. Over half (63.6%) of the caregivers had a general equivalency diploma (GED) or high school diploma or less. Just over one third (34.9%) of the caregivers were married or living as married, 31.8% were divorced or separated, and 33.3% were single. Roughly two thirds (68.2%) of caregivers were unemployed, and 56.1% reported an annual household income of < \$9,999; an additional 16.7% had an income of < \$20,000. Nearly all children (97.0%) were directly exposed to the hurricane (i.e., they were in a building that was damaged, were injured, or were in a community that was flooded; a family member was injured or killed; or they otherwise experienced a loss) and, as a result, 95.7% of the families had to evacuate their communities and 90.0% became homeless. In the year following the hurricane, families moved an average of 3.1 times ($SD = 2.1$, range = 0–11). At T1, 36.4% of participants were living in FEMA trailers, 16.7% in their own house or apartment, 37.9% in a rented apartment or house, 6.1% lived with relatives, and 2.9% reported other living arrangements.

Measures

Children completed the following measures at T1, unless otherwise noted.

Hurricane-related exposure. Children completed an eight-item checklist that parallels in content the Hurricane-Related Exposure scale (La Greca et al., 1996; e.g., "You were injured or hurt"). A total score of events reported was computed.

Child PTSS. The UCLA PTSD Reaction Index (UCLA PTSD–R1; Steinberg, Brymer, Decker, & Pynoos, 2004) assessed PTSS at T1 and T2. The scale reflects *DSM–IV* (APA, 1994) PTSD-related symptoms: reexperiencing (i.e., intrusive memories, nightmares; five items), arousal (i.e., irritability, sleep difficulties; five items), and avoidance (i.e., avoiding people, activities, feelings; seven items). The scale also includes two items assessing other symptoms of clinical significance (i.e., fears of recurrence and trauma-related guilt). Children reported how much of the time they had experienced these symptoms in relation to Hurricane Katrina during the past month on a scale ranging from 0 (*none*) to 4 (*most of the time*). The UCLA PTSD–R1 has demonstrated good internal consistency and test–retest reliability. A PTSS total score was computed, using the method specified by Steinberg et al. (2004); α in this sample = .88.

Posttraumatic growth. The Posttraumatic Growth Inventory for Children–Revised (PTGI-C–R; Kilmer et al., 2009) was administered at T1 and T2. Each youth responded to open-ended items assessing changes perceived in their lives and themselves since Katrina. Children also responded to 10 items assessing changes in five PTG domains: *New Possibilities* ("I have new ideas about how I want things to be when I grow up"), *Relating to Others* ("I feel closer to other people [friends and family] than I used to"), *Personal Strength* ("I learned that I can deal with more things than I thought"), *Appreciation of Life* ("I know what is important to me better than I used to"), and *Spiritual Change* ("My faith [belief] in God is stronger than it was before"; Kilmer et al., 2009). Children responded on a 4-point scale (0 = *no change*, 1 = *a little change*, 2 = *some change*, 3 = *a lot of change*). Consistent with prior work, a total score was computed by summing the items; the scale demonstrated adequate internal reliability at T1 (α = .77) and T2 (α = .81).

Coping competency beliefs. This adaptation of the five-item Children's Competency Beliefs Scale (CCBS; Weyer & Sandler, 1998) used in a study of children displaced by Hurricane Floyd (Cryder et al., 2006) assessed children's perceptions of their ability to handle problems related to the hurricane (e.g., "Overall, I think that the things I did after the hurricanes worked to make things go better for me"), as well as other areas of their life (e.g., "I feel good about the way I have tried to handle my problems"), using a 0 (*not at all true for me*) to 3 (*very true for me*) scale. Mean scores were created; higher scores indicate greater coping competence (α = .63).

Perceived competence. Adapted from Harter's (1982) Self Perception Profile for Children, this 28-item version employs simplified item wording and a single-step forced-choice format (instead of the original's two-stage forced-choice response) to assess self-views of *Social Competence/Acceptance*, *Scholastic Competence*, *Behavioral Conduct*, *Athletic Competence*, *Physical Appearance*, and *Global Self-Worth* (Hoyt-Meyers et al., 1995). For each item, youth indicated which of two statements better described them (e.g., "I don't have as many friends as I'd like" vs. "I have lots of friends"). Items indicating greater perceived competency were coded as 1 and those reflecting lower competence as 0; responses were summed (α = .83).

Future expectations. This seven-item adaptation of Wyman et al.'s (1993) scale assesses youngster's views about their futures (e.g., the degree to which they expect to have a happy life, have friends and people who care about them) on a 3-point scale. The scale was modified to add structure and simplify responses for young children (Kilmer, 1996). Items were summed to create a total score; higher scores suggest more positive future expectations (α = .73).

Realistic control expectations. This adaptation of Wannon's (1990) measure assessed control expectations for controllable (e.g., "can children keep from failing a test if they study?") and uncontrollable (e.g., "can children keep adults from having any worries?") events, using a 3-point scale. This modification retained 11 of 12 items from the scale used by Hoyt-Meyers et al. (1995), dropping one of the two items excluded by Kilmer (1996) to reduce redundancy. Controllable and uncontrollable subscale αs = .70 and .52, respectively; given the low alpha, the latter was not included in analyses. Higher mean scores indicate more accurate expectations.

Rumination Scale for children. Used by Cryder et al. (2006), this five-item adaptation of the adult Rumination Scale (Calhoun, Cann, Tedeschi, & McMillan, 2000) measures intrusive (two items; e.g., "I think about it when I don't mean to") and deliberate rumination, (three items; e.g., "Sometimes I think about it to try to figure out why things like that happen"). Children responded using a scale from 1 (*not at all true for me*) to 4 (*very true for me*); a mean score for each subscale was created to facilitate interpretation. The correlation between the items reflecting intrusive thoughts was .33; the deliberate rumination subscale α = .65.

Caregiver warmth and acceptance. Children completed a 10-item version of a scale used previously to assess perceptions of caregiver warmth and acceptance (e.g., "My caregiver really understands me'' and ''My caregiver enjoys spending time with me''; Greenberger & Chen, 1996). Respondents used a 1 (*not all true*) to 4 (*very true*) scale, and a mean score was created (α = .68).

Caregivers completed the following measures at T1.

Demographic characteristics. Caregivers provided information regarding their age, gender, ethnic background, marital status, employment, annual income, and housing status. In addition, caregivers provided information regarding their child's age, grade, and ethnicity.

Psychological distress. The Hopkins Symptom Checklist–25 (HSCL–25; Derogatis, Lipman, Rickels, Uhlenluth, & Covi, 1974) assessed symptoms of depression, anxiety, and somatization. Caregivers indicated how frequently they had experienced symptoms over the past week using a scale from 0 (*not at all*) to 3 (*extremely*). A mean score was computed (α = .96).

Posttraumatic growth. Caregivers completed the 21-item PTGI (Tedeschi & Calhoun, 1996) at T1. Caregivers indicated to what extent they had experienced changes in their lives and themselves as a result of Katrina and its aftermath on a scale ranging from 0 (*no change*) to 5 (*a great deal of change*); a total summary score was computed. The scale demonstrated excellent internal reliability (α = .94).

Positive reframing coping advice. At T1 caregivers answered two items assessing how frequently they had encouraged their child to use positive reframing to cope postdisaster (e.g., ''encouraged [my child] to see the positive in what happened''; Gil-Rivas et al., 2007). They responded on a 0 (*not at all*) to 4 (*most of the time*) scale. A mean score was created; item r = .34.

Plan of Analysis

Correlations examined associations between key variables. Hierarchical multiple regressions explored correlates of children's PTG scores at T1 by entering caregiver variables in Step 1, child self-system variables and realistic control expectations in Step 2, and intrusive and deliberate rumination in Step 3. At T2, structurally similar regression analyses tested correlates of PTG after controlling for T1 PTG. In an effort to identify the most parsimonious model, variables not associated with PTG at $p < .10$ (on the basis of bivariate correlations) were excluded from the final analyses. As distress has been hypothesized to serve a catalytic role in the PTG process, and PTSS have been shown to be significantly associated with PTG (e.g., Alisic et al., 2008; Kilmer et al., 2009; Salter & Stallard, 2004), similar analyses included concurrent child-reported PTSS to determine the extent to which these symptoms influenced the pattern of relations identified. Finally, as children's perceived coping competency, control expectations, and rumination may vary by age, the following interaction terms were tested: Age × Coping Competency Beliefs, Age × Control Expectations, Age × Intrusive Rumination, and Age × Deliberate Rumination.

Attrition analyses were conducted to determine whether children who did not participate at T2 (n = 15) were significantly different from those who completed the follow-up (n = 51). No significant differences were identified in demographics and the study's core variables.

Results

Table 1 presents descriptive statistics for key study variables. Overall, children reported moderate levels of PTG and mild to moderate levels of PTSS at T1 and T2. Children reported positive future expectations and moderate levels of perceived competence, realistic control expectations, perceived coping competence, and intrusive and deliberate event-related rumination. Overall, children perceived their primary caregiver as warm and

Table 1
Descriptive Statistics for Key Variables of Interest

	M (*SD*)	Scale range
Posttraumatic growth		0–30
Baseline (T1[a])	20.00 (6.54)	
Follow-up (T2[b])	19.22 (7.09)	
Hurricane-related exposure	5.33 (2.63)	0–8
Posttraumatic stress symptoms		0–68
Baseline (T1)	28.23 (15.76)	
Follow-up (T2)	23.98 (12.70)	
Perceived competence	16.80 (5.52)	0–28
Future expectations	17.56 (3.04)	0–21
Coping competency beliefs	1.98 (0.68)	0–3
Realistic control expectations	2.28 (0.57)	1–3
Rumination		1–4
Intrusive	2.37 (1.00)	
Deliberate	2.84 (0.88)	
Perceived caregiver warmth and acceptance	3.29 (0.47)	1–4
Caregiver psychological distress	1.08 (0.79)	0–3
Caregiver posttraumatic growth	71.36 (26.41)	0–105
Positive reframing coping advice	3.41 (0.71)	0–4

[a]Sample size = 66. [b]Sample size = 51.

accepting. Caregivers reported mild levels of distress, moderate levels of PTG, and having provided their child with positive reframing coping advice a good portion of the time.

Associations Among Key Variables

Caregiver and child demographic variables were not significantly associated with children's self-reported PTG at T1 and T2. Table 2 presents correlations among variables of interest. Contrary to expectations, self-system variables, namely perceived competence, future expectations, and coping competency beliefs, did not evidence a significant positive association with PTG at T1. In fact, perceived competence was significantly negatively related to PTG. At T2, among self-system variables, only children's future expectations showed a trend toward a positive association with PTG. Children's realistic control expectations were not significantly associated with T1 or T2 PTG scores. As expected, both intrusive and deliberate rumination were positively associated with children's PTG at T1 and T2. Caregiver warmth was positively associated with children's coping competency beliefs and future expectations, but not with perceived competence or PTG. The provision of coping advice involving positive reframing was the only caregiver variable significantly associated with T1 PTG. Contrary to expectations, no caregiver variable was significantly associated with T2 PTG.

Correlates of PTG

The hierarchical regression analyses revealed that, contrary to expectations, coping advice involving positive reframing was not significantly associated with PTG at T1, after including intrusive and deliberate rumination (see Table 3). Although perceived competence was negatively associated with PTG, it did not reach significance in the final model. Finally, as expected, deliberate rumination was related to higher PTG scores. Analyses not shown found that the inclusion of T1 PTSS did not change the pattern of associations.

At T2, after adjusting for T1 PTG and caregiver distress, although they approached significance ($p = .066$), children's future expectations were not significantly associated with T2 child PTG scores in the final model (Table 4). Moreover, intrusive rumination, rather than deliberate rumination, was associated with higher PTG scores. The inclusion of T2 PTSS minimally altered the results obtained, such that future expectations reached significance in the final step. However, because of the overlap (both conceptual and statistical) between PTSS and intrusive rumination, and in the interest of building the most parsimonious model, T2 PTSS was not included in the final model presented here. Neither age nor the age-related interactions contributed meaningfully to the regressions at either time point.

Discussion

This study sought to extend the research base on PTG in children, examining the role of child self-system and cognitive variables, as well as the caregiving context, in the youngest sample in the PTG literature. These children and their families, impacted by Katrina and its aftermath, were exposed to extreme levels of adversity, and the children evidenced negative consequences 1 year post-disaster (Gil-Rivas et al., 2010). In addition to ongoing distress, a sizable proportion reported PTG (Kilmer et al., 2009). Overall, although conclusions about causality are not possible, results suggest that rumination, both negative, distressing thoughts and deliberate, repetitive thinking, may play a key role in PTG.

Age related neither to PTG nor to the child variables assessed, a result that probably reflects the sample's restricted age range. Given that prior age-related findings have varied (e.g., Cryder et al., 2006; Milam et al., 2004), future studies could examine the degree to which variables hypothesized to be relevant to the PTG process have a differential impact across developmental stages (Kilmer, 2006).

Hypotheses involving self-system functioning, operationalized by perceptions of one's competence, coping competency beliefs, and future expectations, received minimal support. For instance, consistent with hypotheses, positive future expectations were identified as a potentially "active" variable. Specifically, although not meaningfully related to T1 PTG, after adjusting for T1 PTG and caregiver distress, future expectations at T1 approached significance in contributing to the prediction of T2 PTG in the final regression. Although the T2 finding is suggestive, these results involving future expectations are largely not consistent with adult PTG findings regarding the influence of optimism (Linley & Joseph, 2004) or child resilience research (Werner & Smith, 1982; Wyman et al., 1993). Overall, while associations involving this variable demonstrated expected relations with other self-system variables at T1, such that optimistic views about one's future

Table 2

Correlations Among Variables of Interest

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Child PTG T1	—															
2. Child PTG T2	.44**	—														
3. Child age	−.10	.07	—													
4. Hurricane exposure	.10	.20†	.03	—												
5. PTSS T1	.45***	.32*	−.01	.51***	—											
6. PTSS T2	.27*	.32*	−.18†	.30*	.48***	—										
7. Perceived competence	−.24*	−.03	−.17†	−.32**	−.39**	−.10	—									
8. Coping competency beliefs	−.02	.08	.04	−.03	.11	−.20†	.21*	—								
9. Future expectations	.15	.22†	−.04	−.15	−.21*	−.12	.41***	.33**	—							
10. Realistic control expectations	−.16	−.18	.01	−.05	−.25*	−.18	.04	.10	.07	—						
11. Rumination—intrusive	.51***	.46***	−.01	.32**	.70***	.35**	−.35**	−.08	−.02	−.28*	—					
12. Rumination—deliberate	.54***	.35**	−.01	.33**	.59***	.39**	−.12	.07	.20†	−.23*	.64***	—				
13. Perceived caregiver warmth	.14	.13	.26*	−.16	−.09	−.22†	.14	.30*	.32**	−.02	.04	.25*	—			
14. Caregiver PTG T1	.16	.04	.05	.25*	.36**	.38**	−.29**	−.15	−.14	−.13	.25*	.13	−.17†	—		
15. Coping advice—positive reframing	.30**	.16	−.07	.05	.18†	.15	.08	−.14	−.01	−.22*	.17†	.35**	.02	.45***	—	
16. Caregiver distress T1	.03	.19†	.05	.25*	.26*	.25*	−.17†	−.02	−.07	−.19†	.10	.04	−.10	.26*	−.01	—

Note. T1: $n = 66$; T2: $n = 51$. PTG = posttraumatic growth; PTSS = posttraumatic stress symptoms.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 3
Summary of Hierarchical Regression Analysis for Child and Caregiving Variables Predicting Posttraumatic Growth at Baseline

	Step 1		Step 2		Step 3	
	B	β	B	β	B	β
Coping advice—positive reframing	2.73	.30*	2.91	.32**	1.43	.16
Perceived competence	—	—	−0.31	−.27*	−0.16	−.13
Rumination—intrusive	—	—	—	—	1.53	.23
Rumination—deliberate	—	—	—	—	2.39	.32*

Note. Step 1: Adjusted $R^2 = .07$, $p = .015$; Step 2: Adjusted $R^2 = .13$, $p = .004$; Step 3: Adjusted $R^2 = .33$, $p = .001$.
$^{*}p < .05$. $^{**}p < .01$.

Table 4
Summary of Hierarchical Regression Analysis for Baseline Child and Caregiving Variables Predicting Posttraumatic Growth at Follow-Up

	Step 1		Step 2		Step 3	
	B	β	B	β	B	β
Posttraumatic growth T1	0.46	.42**	0.43	.40**	0.24	.22
Caregiver distress T1	1.54	.16	1.78	.18	1.54	.16
Future expectations	—	—	0.47	.19	0.58	.24†
Rumination—intrusive	—	—	—	—	2.48	.35*
Rumination—deliberate	—	—	—	—	−0.09	−.01

Note. Step 1: Adjusted $R^2 = .18$, $p = .003$; Step 2: Adjusted $R^2 = .20$, $p = .003$; Step 3: Adjusted $R^2 = .26$, $p = .002$.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

correlated with perceived competence and coping competency beliefs, it is not possible to ascertain (a) whether these expectations anteceded or were elicited by the other self-system competencies assessed, or (b) of more salience here, their specific role in the PTG process. Wyman et al. (1993) found that positive future expectations reported by 9- to 11-year-olds seemingly promoted ongoing positive adaptation, both directly and by moderating the adverse effects of subsequent stress exposure. However, following a trauma of Katrina's magnitude, as well as the persistent secondary adversities experienced by families, it may have been difficult for the children to maintain hope and optimism about their future. Nevertheless, it is possible that such future views may facilitate not only positive adaptation but also growth. That is, hope for the future may provide a lens through which children can try to view, cope with, and make sense of their circumstance, perhaps increasing the likelihood that they will try to identify the positive in their situation. Future research can address their potential role in the PTG process.

The hypothesis that the study's other self-system variables would relate positively to PTG was not supported. Specifically, in contrast to expectations and results linking competency beliefs and PTG reported by Cryder et al. (2006), a study that also included older children, coping competency beliefs did not relate to PTG at T1 or T2. Perhaps more notably, perceived competence related negatively to PTG at T1 (both at the bivariate level and in the regression, although not in the final model) and did not contribute to predictions of PTG at T2. Variables reflecting perceived competence have discriminated sensitively between those evidencing resilient adaptation and those experiencing adjustment problems (Hoyt-Meyers et al., 1995). The present findings may thus point to another point of divergence in the resilience and PTG processes. That is, the same self-views of competence that seemingly help forestall dysfunction in the face of adversity may very well reduce the likelihood of growth posttrauma. Put another way, perceptions of one's competence may foster positive adjustment, but not PTG. If one holds positive views of his or her competence, it is possible that he or she may appraise the event and its impact for him- or herself and his or her world differently; as such, these children may not perceive their assumptive worlds as "shattered" to the same degree. In turn, it may be that those children who reported more positive self-views regarding their competence may experience less distress and reduced levels of intrusive ideation and, consequently, given the presumptive link between distress and PTG, not go through a growth-like process. This notion is supported by the significant negative correlations between perceived competence and both intrusive rumination and T1 PTSS in this sample. On the other hand, if one is experiencing high levels of PTSS, as well as intrusive rumination, perhaps one path to adaptation is via deliberate, constructive rumination and PTG. The PTG evidence base has not yet established if those reporting PTG also evidence proportionally fewer PTSS over time (e.g., Salter & Stallard, 2004); however, this will be a consideration for future longitudinal examinations. Alternatively, it may be that the self-system, more globally, and competency beliefs, in particular, become more salient as children progress along their developmental trajectories and their self-views crystallize, while other factors, such as the role of the caregiver, carry special weight for younger children.

Indeed, future research will need to elucidate the role of caregivers in fostering PTG in children, because the present findings are mixed. Although caregiver warmth correlated positively with two of the study's three self-system variables (coping competency and future expectations), it did not relate to PTG. Although unexpected here, this finding is consistent with the results of Cryder et al. (2006). In considering these findings, it may be that, analogous to the results involving perceived competence, caregiver warmth and responsiveness may serve a protective function under adverse conditions (Luthar et al., 2000; Masten & Coatsworth, 1998) but may reduce the likelihood of PTG. Warm, supportive caregiving may affect youngsters' appraisal of the trauma and reduce the degree to which they feel shaken in its aftermath, thereby limiting the experience of ongoing distress thought so critical to the PTG process.

That caregiver distress did not relate significantly to the core child variables was also unexpected (though the correlation with T2 PTG approached significance). This factor is presumed to relate to availability and the capacity to provide support and nurturance, even coping advice. The present data do not convey the degree to which children were aware of caregiver distress; it may be that this factor holds salience when children perceive and understand caregivers' reactions and difficulties. This notion may also explain the lack of association between caregiver and child PTG. That is, if children are unaware of positive changes experienced by caregivers, they may not be influenced by them. It may be that such positive changes—or the orientation to seek the positive in a difficult situation—manifest in the coping guidance offered by caregivers. Notably, among the caregiver variables, only positive reframing coping advice was associated with PTG at T1, although it did not reach significance in the final model; it also did not relate to T2 PTG. The process of encouraging one's child(ren) to identify the positive in their new circumstance would appear to be of direct conceptual relevance to PTG, and the present data—including positive associations with deliberate rumination and caregiver PTG—are suggestive. The role of this variable warrants attention in future research.

Realistic control, thought to be a key cognitive resource for children, also did not function as expected. In fact, this slightly modified version of a scale used by Hoyt-Meyers et al. (1995) did not hold up psychometrically in the present sample, leading to the exclusion of the subscale for uncontrollable events. While its modification may have contributed to the scale's reduced psychometric properties, this outcome was unexpected, particularly since, as Wyman (2003) noted, the Hoyt-Meyers study demonstrated that many children in their young cohort (aged 7–9 years) had the cognitive capacity to report differentiations regarding the controllability of the events assessed. That this measure performed so poorly raises the possibility that, given the youngsters' experiences and belief systems, and the values and practices of their culture(s), the items on these scales took on a different meaning in this postdisaster context. Beyond that, the controllable subscale correlated negatively with intrusive rumination and T1 PTSS, indicating that more accurate views about controllable events tended to be associated with lower levels of intrusive thinking and PTSS. Nevertheless, realistic control did not contribute reliably to prediction models. Future research will have to examine the role of this construct in PTG.

The most noteworthy findings center on rumination. Consistent with predictions, T1 deliberate rumination related positively to PTG, after accounting for caregiving and self-system variables. In fact, this variable was the lone significant predictor of PTG in the final T1 model. In addition, T1 levels of intrusive rumination, but not deliberate rumination, significantly predicted PTG at T2, providing partial support for the hypothesized role of rumination. Although T1 levels of deliberate rumination correlated with PTG at T2 ($r = .35$), this variable did not add meaningfully to the regression model. Overall, in line with PTG theory, variables reflecting ruminative processing exhibited strong relations to growth. The present findings highlight the two complementary cognitive processes hypothesized to be at play: (a) negative, distressing ideation associated with one's ongoing struggle with adversity and his or her new reality seemingly catalyzes the process leading to growth, and (b) deliberate rumination serves a constructive function, as one works to reprocess the trauma and its aftermath, in an effort to understand and reconcile the experience with prior models of self, others, and the world. The present design cannot test these notions directly, but its results are suggestive.

That ruminative thinking was directly associated with PTG is consistent with findings reported in adult research (Linley & Joseph, 2004), but does not accord with the findings of Cryder et al. (2006)—these researchers did not examine intrusive and deliberate rumination separately and did not detect an association between rumination and PTG; rather, rumination related significantly to competency beliefs, which were associated with PTG. This study's hypotheses regarding the weight of rumination were informed by assumptions about the importance of cognitive reprocessing of the trauma to PTG, as well as the still-emerging representational worlds and self-systems of young children. It may be that the roles of self-system and rumination variables vary by age. This possibility, and the nature of the rumination–PTG relation, will need to be clarified in subsequent work. Given developmental considerations in the study of rumination (see Sprung, 2008), future studies should examine these processes in relation to PTG, in particular the constructive, cognitive reprocessing of one's circumstances, in larger, more diverse samples.

Study Limitations, Contributions, and Future Directions

Several study limitations bear mention. The study's relatively small sample affected power, prohibited the use of other statistical techniques, and limited foci to the variables selected. In addition, the sample's voluntary nature limits generalizability because nonrandom factors may have influenced families' decisions to participate. Moreover, sample characteristics (i.e., evacuees, residents of FEMA-provided housing) also limit generalizability. Other aspects of the sample's composition and cultural context may also have impacted results. For instance, the sample largely includes African American children from impoverished backgrounds, many of whom may have had significant prior trauma exposure (e.g., Osofsky et al., 2007), factors that may have influenced their post-Katrina response (e.g., appraisals and attributions), capacity (or willingness) to access services and supports, and adaptation. In addition, many evacuee children (likely including study respondents) had been attending underperforming schools and/or had been assessed as performing below national grade-level norms (Osofsky et al., 2007; Snider, Hoffman, Littrell, Fry, & Thornburgh, 2010); these factors are relevant in light of the cognitive components of the hypothesized PTG process and may influence generalizability. These results highlight future directions for child PTG research, including work investigating the function of context (Kilmer et al., 2009). Finally, that T1 was 1 year postdisaster restricts the degree to which one can draw clear temporal inferences regarding the roles of the factors associated with PTG. That is, it is unclear if they were resources or approaches present pretrauma versus "products" that increased, emerged, or evolved in the trauma's aftermath. Prospective-longitudinal designs are necessary to enhance understanding of the PTG process.

Nevertheless, the child PTG literature is nascent in its development, and the present findings contribute to the knowledge base in an area in which, to date, there has been a paucity of empirical work. Notwithstanding the limitations described, this work provides an examination of child and caregiving context variables among young children who had been exposed to severe trauma and highlights key correlates of PTG. In the early years of a developing area, it is critical to identify factors associated with the construct or process of interest. Such generative work can then be built upon with the use of methodologies that allow for exploration of the pathways seemingly involved in the growth process. The present findings demonstrate the merit of studying variables reflecting child resources and processes (e.g., rumination) and the caregiving environment in the context of PTG work and serve to identify factors that might profitably be incorporated into designs in which causal inferences are possible.

In addition, because the domains in which children evidence growth may vary with age, future efforts could compare the nature of PTG across different stages, that is, the degree to which there is a developmental progression in the manner in which PTG manifests. It may be that young children can more easily experience and express growth around their perceptions of their strength, their relationships, and their spirituality (particularly if caregivers have invoked faith-based explanations and/or encouraged faith-based coping), rather than describing a changed philosophy of life or a sense of new possibilities. Some of these presumed differences stem from the cognitive demands and self-awareness necessary to experience PTG, as well as the fact that children's sense of self is evolving and self-schema are less crystallized, thereby likely limiting the degree to which young children will describe growth in these areas. Future research could also explore the degree to which the PTG process (or one's growth trajectory) may differ following acute versus chronic traumas, or traumas with significant postevent adversities. Efforts investigating such questions could yield findings of relevance to those working in trauma, developmental psychology, and developmental psychopathology.

In addition to its heuristic benefit, research enhancing understanding of the PTG construct and the growth process among young people can yield valuable information for those working with youth who have experienced trauma (Kilmer, 2006). For example, although the present findings must be considered preliminary, they suggest the potential benefit of working to restructure youngsters' cognitive appraisals of trauma and foster productive rumination. Work in this area can also increase the likelihood that clinicians and other professionals will attend to and assess relevant competencies and resources, and take steps to facilitate resources and processes that may foster PTG (Kilmer & Gil-Rivas, 2008; Tedeschi & Kilmer, 2005).

References

Aldwin, C. M. (2007). *Stress, coping, and development: An integrative perspective* (2nd ed.). New York: Guilford.

Alisic, E., van der Schoot, T. A. W., van Ginkel, J. R., & Kleber, R. J. (2008). Looking beyond posttraumatic stress disorder in children: Posttraumatic stress reactions, posttraumatic growth, and quality of life in a general population sample. *Journal of Clinical Psychiatry, 29*, 1455–1461.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Barakat, L. P., Alderfer, M. A., & Kazak, A. E. (2006). Posttraumatic growth in adolescent survivors of cancer and their mothers and fathers. *Journal of Pediatric Psychology, 31*, 413–419.

Calhoun, L. G., Cann, A., Tedeschi, R. G., & McMillan, J. (2000). A correlational test of the relationship between posttraumatic growth, religion, and cognitive processing. *Journal of Traumatic Stress, 13*, 521–527.

Calhoun, L. G., & Tedeschi, R. G. (2006). The foundations of posttraumatic growth: An expanded framework. In L. G. Calhoun & R. G. Tedeschi (Eds.), *Handbook of posttraumatic growth: Research and practice* (pp. 1–23). Mahwah, NJ: Erlbaum.

Cicchetti, D. (2003). Foreword. In S. Luthar (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. xix–xxvii). New York: Cambridge University Press.

Cicchetti, D., & Lynch, M. (1993). Toward an ecological/transactional model of community violence and child maltreatment: Consequences for children's development. *Psychiatry, 56*, 96–118.

Cohen, L. H., Hettler, T. R., & Pane, N. (1998). Assessment of posttraumatic growth. In R. G. Tedeschi, C. L. Park, & L. G. Calhoun (Eds.), *Posttraumatic growth: Positive change in the aftermath of crisis* (pp. 23–42). Mahwah, NJ: Erlbaum.

Compas, B. E., Connor-Smith, J., Saltzman, H., Harding Thomsen, A., & Wadsworth, M. E. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin, 127*, 87–127.

Cowen, E. L., Work, W. C., Wyman, P. A., Parker, G. R., Wannon, M., & Gribble, P. A. (1992). Test comparisons among stress-affected, stress-resilient and nonclassified fourth through sixth grade urban children. *Journal of Community Psychology, 20*, 200–214.

Cryder, C. H., Kilmer, R. P., Tedeschi, R. G., & Calhoun, L. G. (2006). An exploratory study of posttraumatic growth in children following a natural disaster. *American Journal of Orthopsychiatry, 76*, 65–69.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenluth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist: A self-report symptom inventory. *Behavioral Science, 19*, 1–15.

Fritz, G. K., Williams, J. R., & Amylon, J. R. (1988). After treatment ends: Psychosocial sequelae in pediatric cancer survivors. *American Journal of Orthopsychiatry, 58*, 552–561.

Gil-Rivas, V., Holman, E. A., & Silver, R. C. (2004). Adolescent vulnerability following the September 11th terrorist attacks: A study of parents and their children. *Applied Developmental Science, 8*, 130–142.

Gil-Rivas, V., Kilmer, R. P., Hypes, A. W., & Roof, K. A. (2010). The caregiver-child relationship and children's adjustment post-Hurricane Katrina. In R. P. Kilmer, V. Gil-Rivas, R. G. Tedeschi, & L. G. Calhoun (Eds.), *Helping families and communities recover from disaster: Lessons learned from Hurricane Katrina and its aftermath.* (pp. 55–76) Washington, DC: American Psychological Association.

Gil-Rivas, V., Silver, R. C., Holman, E. A., McIntosh, D., & Poulin, M. (2007). Parental response and adolescent adjustment to the September 11, 2001 terrorist attacks. *Journal of Traumatic Stress, 20*, 1063–1068.

Grant, K. E., Compas, B. E., Stuhlmacher, A. F., Thurm, A. E., McMahon, S. D., & Halpert, J. A. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological Bulletin, 129*, 447–466.

Greenberger, E., & Chen, C. (1996). Perceived family relationships and depressed mood in early and late adolescence: A comparison of European and Asian Americans. *Developmental Psychology, 32*, 707–716.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87–97.

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition, 4*, 119–151.

Harter, S. (2006). Self-processes and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 1. Theory and method* (2nd ed., pp. 370–418). Hoboken, NJ: John Wiley & Sons.

Harter, S., & Buddin, B. (1987). Children's understanding of the simultaneity of two emotions: A five stage developmental acquisition sequence. *Developmental Psychology, 23*, 388–399.

Hasan, N., & Power, T. G. (2004). Children's appraisal of major life events. *American Journal of Orthopsychiatry, 74*, 26–32.

Helgeson, V. S., Reynolds, K. A., & Tomich, P. L. (2006). A meta-analytic review of benefit finding and growth. *Journal of Consulting and Clinical Psychology, 74*, 797–816.

Hoyt-Meyers, L. A., Cowen, E. L., Work, W. C., Wyman, P. A., Magnus, K. B., Fagen, D. B., et al. (1995). Test correlates of resilient outcomes among highly stressed 2nd and 3rd grade urban children. *Journal of Community Psychology, 23*, 326–338.

Hurricane Katrina Community Advisory Group. (2006). Overview of baseline survey results. Retrieved November 22, 2007, from http://hurricanekatrina.med.harvard.edu/pdf/baseline_report%208-25-06.pdf

Ickovics, J. R., Meade, C. S., Kershaw, T. S., Milan, S., Lewis, J. B., & Ethier, K. A. (2006). Urban teens: Trauma, posttraumatic growth, and emotional distress among female adolescents. *Journal of Consulting and Clinical Psychology, 74*, 841–850.

Janoff-Bulman, R. (1992). *Shattered assumptions*. New York: Free Press.

Kilmer, R. P. (1996, October). *Pre-post evaluation of a school mental health program: A pilot study using child measures.* Poster session presented at the annual convention of the New York Association of School Psychologists, Rochester, NY.

Kilmer, R. P. (2006). Resilience and posttraumatic growth in children. In L. G. Calhoun & R. G. Tedeschi (Eds.), *Handbook of posttraumatic growth: Research and practice* (pp. 264–288). Mahwah, NJ: Erlbaum.

Kilmer, R. P., & Gil-Rivas, V. (2008). Posttraumatic growth in youth following disasters. *The Prevention Researcher, 15*, 18–20.

Kilmer, R. P., & Gil-Rivas, V. (2010). Introduction: Attending to ecology. In R. P. Kilmer, V. Gil-Rivas, R. G. Tedeschi, & L. G. Calhoun (Eds.), *Helping families and communities recover from disaster: Lessons learned from Hurricane Katrina and its aftermath* (pp. 3–24). Washington, DC: American Psychological Association.

Kilmer, R. P., Gil-Rivas, V., Tedeschi, R. G., Cann, A., Calhoun, L. G., Buchanan, T., et al. (2009). Use of the revised Posttraumatic Growth Inventory for Children (PTGI-C-R). *Journal of Traumatic Stress, 22*, 248–253.

Kliewer, W., Sandler, I., & Wolchik, S. (1994). Family socialization of threat appraisal and coping: Coaching, modeling, and family context. In F. Nestmann & K. Hurrelmann (Eds.), *Social networks and social support in childhood and adolescence* (pp. 271–291). New York: Waler de Gruyter.

Krementz, J. (1983). *How it feels when a parent dies*. New York: Knopf.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (1996). Symptoms of posttraumatic stress in children following Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology, 64*, 712–723.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (Eds.). (2002). *Helping children cope with disasters and terrorism*. Washington, DC: American Psychological Association.

Laufer, A., & Solomon, Z. (2006). Posttraumatic symptoms and posttraumatic growth among Israeli youth exposed to terror incidents. *Journal of Social and Clinical Psychology, 25*, 429–447.

Lieberman, A. F., & Van Horn, P. (2004). Assessment and treatment of young children exposed to traumatic events. In J. D. Osofsky (Ed.), *Young children and trauma: Intervention and treatment* (pp. 111–138). New York: Guilford.

Linley, P. A., & Joseph, S. (2004). Positive change following trauma and adversity: A review. *Journal of Traumatic Stress, 17*, 11–21.

Luthar, S. S. (Ed.). (2003). *Resilience and vulnerability: Adaptation in the context of childhood adversities*. New York: Cambridge University Press.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*, 543–562.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*, 227–238.

Masten, A. S., Best, K. M., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*, 425–444.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments. *American Psychologist, 53*, 205–220.

Milam, J. E., Ritt-Olson, A., & Unger, J. (2004). Posttraumatic growth among adolescents. *Journal of Adolescent Research, 19*, 192–204.

National Child Traumatic Stress Network. (n.d.). *Age-related reactions to a traumatic event*. Retrieved April 9, 2009, from http://www.nctsn.org/nctsn_assets/pdfs/age_related_reactions.pdf

Osofsky, J. D. (Ed.). (2004a). *Young children and trauma: Intervention and treatment*. New York: Guilford.

Osofsky, J. D. (2004b). Different ways of understanding young children and trauma. In J. D. Osofsky (Ed.), *Young children and trauma: Intervention and treatment* (pp. 3–9). New York: Guilford.

Osofsky, J. D., Osofsky, H. J., & Harris, W. W. (2007). Katrina's children: Social policy considerations for children in disasters. *Society for Research in Child Development Social Policy Report, 21*, 1–18.

Osofsky, J., Osofsky, H. J., Kronenberg, M., & Cross, T. (2010). The aftermath of Hurricane Katrina: Mental health considerations and lessons learned. In R. P. Kilmer, V. Gil-Rivas, R. G. Tedeschi, & L. G. Calhoun (Eds.), *Helping families and communities recover from disaster: Lessons learned from Hurricane Katrina and its aftermath* (pp. 241–263). Washington, DC: American Psychological Association.

Pfefferbaum, B., Nixon, S. J., Tucker, P. M., Tivis, R. D., Moore, V. L., Gurwitch, R. H., et al. (1999). Posttraumatic stress responses in bereaved children after the

Oklahoma City bombing. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1372–1379.

Phipps, S., Long, A., & Ogden, J. (2007). Benefit finding scale for children: Preliminary findings from a childhood cancer population. *Journal of Pediatric Psychology, 32*, 1264–1271.

Pynoos, R. S., Steinberg, A. M., & Wraith, R. (1995). A developmental model of childhood traumatic stress. In D. J. Cohen & D. Cicchetti (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation* (pp. 72–95). Oxford, England: Wiley.

Salmon, K., & Bryant, R. A. (2002). Posttraumatic stress disorder in children: The influence of developmental factors. *Clinical Psychology Review, 22*, 163–188.

Salter, E., & Stallard, P. (2004). Posttraumatic growth in child survivors of a road traffic accident. *Journal of Traumatic Stress, 17*, 335–340.

Shahinfar, A., & Fox, N. A. (1997). The effects of trauma on children: Conceptual and methodological issues. In D. Cicchetti & S. L. Toth (Eds.), *Rochester symposium on developmental psychopathology: Vol. 8. Developmental perspectives on trauma: Theory, research, and intervention* (pp. 115–139). Rochester, NY: University of Rochester Press.

Silverman, W. K., & La Greca, A. M. (2002). Children experiencing disasters: Definitions, reactions, and predictors of outcomes. In A. M. La Greca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 11–33). Washington, DC: American Psychological Association.

Snider, L., Hoffman, Y., Littrell, M., Fry, W., & Thornburgh, M. (2010). Supporting children post-Katrina: Reflections on psychosocial interventions and guiding principles. In R. P. Kilmer, V. Gil-Rivas, R. G. Tedeschi, & L. G. Calhoun (Eds.), *Helping families and communities recover from disaster: Lessons learned from Hurricane Katrina and its aftermath* (pp. 25–51). Washington, DC: American Psychological Association.

Sprung, M. (2008). Unwanted intrusive thoughts and cognitive functioning in kindergarten and young elementary school-age children following Hurricane Katrina. *Journal of Clinical Child and Adolescent Psychology, 37*, 575–587.

Steinberg, A. M., Brymer, M. J., Decker, K. B., & Pynoos, R. S. (2004). The University of California at Los Angeles Post-Traumatic Stress Disorder Reaction Index. *Current Psychiatry Reports, 6*, 96–100.

Taku, K., Calhoun, L. G., Kilmer, R. P., & Tedeschi, R. G. (2008, March). *Posttraumatic growth and non-traumatic growth in Japanese youth*. Paper session presented at the annual meeting of the Southeastern Psychological Association, Charlotte, NC.

Tedeschi, R. G., & Calhoun, L. G. (1996). The Posttraumatic Growth Inventory: Measuring the positive legacy of trauma. *Journal of Traumatic Stress, 9*, 455–471.

Tedeschi, R. G., Calhoun, L. G., & Cann, A. (2007). Evaluating resource gain: Understanding and misunderstanding posttraumatic growth. *Applied Psychology: An International Review, 56*, 396–406.

Tedeschi, R. G., & Kilmer, R. P. (2005). Assessing strengths, resilience, and growth to guide clinical interventions. *Professional Psychology: Research and Practice, 36*, 230–237.

Wannon, M. (1990). *Children's control beliefs about controllable and uncontrollable events: Their relationship to stress resilience and psychosocial adjustment*. Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*, 163–206.

Weems, C. F., Pina, A. A., Costa, N. M., Watts, S. E., Taylor, L. K., & Cannon, M. F. (2007). Predisaster trait anxiety and negative affect predict posttraumatic stress in youths after Hurricane Katrina. *Journal of Consulting and Clinical Psychology, 75*, 154–159.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible: A study of resilient children*. New York: McGraw-Hill.

Weyer, M., & Sandler, I. N. (1998). Stress and coping as predictors of children's divorce-related ruminations. *Journal of Clinical Child Psychology, 27*, 78–86.

Wyman, P. A. (2003). Emerging perspectives on context specificity of children's adaptation and resilience: Evidence from a decade of research with urban children in adversity. In S. S. Luthar (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. 293–317). New York: Cambridge University Press.

Wyman, P. A., Cowen, E. L., Work, W. C., & Kerley, J. H. (1993). The role of children's future expectations in self-system functioning and adjustment to life stress: A prospective study of urban at-risk children. *Development and Psychopathology, 5*, 649–661.

Yates, T. M., Egeland, B., & Sroufe, L. A. (2003). Rethinking resilience: A developmental process perspective. In S. S. Luthar (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. 243–266). New York: Cambridge University Press.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1228–1240

Sex Differences in Salivary Cortisol, Alpha-Amylase, and Psychological Functioning Following Hurricane Katrina

Jacob M. Vigil
University of New Mexico

David C. Geary
University of Missouri

Douglas A. Granger
The Pennsylvania State University

Mark V. Flinn
University of Missouri

The study examines group and individual differences in psychological functioning and hypothalamic–pituitary–adrenal and sympathetic nervous system (SNS) activity among adolescents displaced by Hurricane Katrina and living in a U.S. government relocation camp ($n = 62$, ages 12–19 years) 2 months postdisaster. Levels of salivary cortisol, salivary alpha-amylase, depression, anxiety, distress, aggression, and self-esteem for this group were contrasted with a demographically matched no-trauma control group ($n = 53$). Results revealed that hurricane exposure and SNS activity moderated the relations between lower cortisol and higher internalizing behaviors. Sex-related differences were observed in behavioral adjustment and stress regulation. Implications of sex differences in biobehavioral adjustment to loss, displacement, and relocation are discussed in relation to evolutionary and developmental theory.

In all cultures in which it has been studied, exposure to and social and material losses resulting from natural disasters is associated with increased prevalence of severe psychological distress (e.g., Chae, Kim, Rhee, & Henderson, 2005; Weinstein, Lyon, Rothman, & Cuite, 2000). Patterns of physiological and psychological reactions to natural disasters, however, may vary with the severity and duration of the attendant losses and by normative sex-related differences in biological, social, and behavioral concomitants of the stress response (Geary & Flinn, 2002; Taylor et al., 2000; Vigil, 2009). In both sexes, short-term physiological responses to acute stressors include (a) increased activation of the hypothalamic–pituitary–adrenal (HPA) axis, and the synthesis and secretion of glucocorticoids (i.e., cortisol) and (b) activation of the sympathetic nervous system (SNS) and release of catecholamines (i.e., norepinephrine, epinephrine) into the bloodstream (e.g., Chrousos & Gold, 1992). In contrast, repeated and chronic stress exposure is associated with low or blunted HPA activity (e.g., low cortisol levels or flat-shallow diurnal pattern of cortisol production), potentially reflecting habituation or adaptation to these circumstances and the overall dampening of HPA reactivity (e.g., Burke, Fernald, Gertler, & Adler, 2005; Flinn, Quinlan, Turner, Decker, & England, 1996; Miller, Chen, & Zhou, 2007).

Contemporary theorists speculate that individual differences in HPA axis activity may play a role in why exposure to adverse events results in major psychological disturbances for some individuals but has minimal consequences for others (e.g., Ellis, Jackson, & Boyce, 2006). Individual differences in HPA activity are correlated with internalizing problem behaviors (e.g., low mood), low levels of control-related beliefs, and anxiety during adolescence (e.g., Granger, Weisz, McCracken, Ikeda, & Douglas, 1996; Klimes-Dougan, Hastings, Granger, Usher, & Zahn-Waxler, 2001). These patterns can sometimes result in prolonged symptoms of posttraumatic stress disorder (PTSD) and major depression disorder, specifically, disassociation, emotional numbness, intrusive and distressful thoughts, and feelings of hopelessness, fatigue, idiopathic (cause unknown) pain, and negative self-evaluations (e.g., Ehlert, Gaab, & Heinrichs, 2001). Repeated activation of HPA and SNS has also been linked to higher morbidity, including changes that may induce

---

We thank Patricia Vigil for her assistance with data collection.

Correspondence concerning this article should be addressed to Jacob M. Vigil, Department of Psychology, University of New Mexico, 1 University of New Mexico, MSC03 2220, Albuquerque, NM 87131-1161. Electronic mail may be sent to vigilj@unm.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0014

lower immune functioning, increased susceptibility to viral infection, and actual onset of certain diseases (e.g., pulmonary constriction; Cohen et al., 1998; Flinn & England, 2003; Kunz-Ebrecht, Mohamed-Ali, Feldman, Kirschbaum, & Steptoe, 2003; Misra et al., 2004; Rotton & Dubitsky, 2002; Segerstrom & Miller, 2004; Uchino, Cacioppo, & Kiecolt-Glaser, 1996).

Recent laboratory studies show that relations between HPA activity and some health-related consequences such as mood disturbances may in turn be moderated by SNS (e.g., El-Sheikh, Erath, Buckhalt, Granger, & Mize, 2008; Gordis, Granger, Susman, & Trickett, 2006). Bauer, Quas, and Boyce (2002) proposed that the influences of the SNS and HPA axis on child adjustment may be conditional upon one another and put forth two testable hypotheses (Bauer et al., 2002). The first hypothesis is based on an "additive" model that assumes HPA activity primarily augments (i.e., operates in parallel with) the SNS, such that symmetrical activity across these systems could result in hypoarousal (low SNS and HPA activity) or hyperarousal (high SNS and HPA activity), and prolonged psychophysiological changes in either direction may leave children susceptible to adjustment problems. Conversely, children would be at least risk for adjustment problems when both systems are only moderately active, or when the activity of the SNS and HPA is asymmetrical (i.e., one is high and the other is low). The second hypothesis is based on an "interactive" model that rests on the possibility that HPA and SNS have different functions, and HPA activity, for example, may manifest differently (e.g., behaviorally) or perhaps suppress (counteract) SNS activity. Based on this hypothesis, HPA may modulate the SNS, and asymmetrical activation of these systems would be associated with more pronounced distress behaviors (e.g., depression and/or aggression), and hence greater psychological impairment. Both models assume that homeostatic (medium) levels of arousal are optimal for wellness, and indeed numerous studies have shown that both low and high arousal levels are associated with adjustment problems (e.g., Kagan, Snidman, Arcus, & Reznick, 1994). It is important to note that both additive and interactive models, as proposed by Bauer et al. (2002), are compatible with *statistical interactions* between the systems; that is, the association between the activity of one system and child adjustment is conditional upon activity of the other system.

Gordis et al. (2006) presented the first empirical investigation of the interaction between HPA (cortisol) and salivary alpha-amylase (sAA), which is a surrogate marker of SNS reactivity (Granger, Kivlighan, El-Sheikh, Gordis, & Stroud, 2007; Rohleder, Nater, Wolf, Ehlert, & Kirschbaum, 2004), as a predictor of aggressive behavior in a sample of early adolescents. Results provided partial support for the "additive" model. Specifically, at low levels of sAA activity, lower cortisol levels were associated with higher parent ratings of aggression, but at high levels of sAA activity, lower cortisol levels were not associated with aggression (Gordis et al., 2006). That is, the highest levels of aggression were found among adolescents with low levels of both cortisol and sAA activity (i.e., symmetrical activity). More recently, El-Sheikh et al. (2008) reported interactions between basal SNS (i.e., as measured by skin conductance and sAA) and cortisol levels, and variance in children's externalizing and internalizing problems (El-Sheikh et al., 2008). Higher basal cortisol levels were positively associated with higher internalizing and externalizing problems among children with higher SNS activity, compared to children with lower SNS activity. In contrast, a recent study among toddlers found partial support for the interactive or specialization model such that sAA was more predictive of positive affect and approach behaviors, whereas cortisol was more predictive of low mood and withdrawal behaviors (Fortunato, Dribin, Granger, & Buss, 2008). These studies add to an emerging body of research that shows sAA and cortisol provide unique information in the prediction of individual differences in behavioral adjustment to stress of children (e.g., El-Sheikh et al., 2008; Gordis, Granger, Susman, & Trickett, 2008; Nater et al., 2006; Stroud et al., 2009). Moreover, this research suggests that inclusion of information about the coordination between HPA and SNS activity may be particularly important for biosocial models of risk and resilience in development science.

All studies to date, however, have only explored sAA and cortisol in relation to problem behavior using acute stress reactivity paradigms among nontraumatized samples in laboratory settings (but see research on maltreated children; Gordis et al., 2008). The present study is the first to examine coordination of HPA and SNS among displaced adolescents that have experienced chronic and pervasive stress caused by high levels of trauma, loss, and relocation. Specifically, we examined psychological adjustment and stress-related psychobiology of adolescents and young adults who lost their homes as a result of Hurricane Katrina, 2 months after the hurricane and several weeks after relocation to a large, semipermanent, U.S. government-provided housing camp. These individuals experienced multiple forms of severe and uncontrollable trauma, including events (e.g., loss of property) that resulted from the hurricane and flooding, as well as the hardship and uncertainty associated with relocating to an unfamiliar community and living under postdisaster conditions that present a unique suite of ongoing stressors (e.g., inadequate housing facilities) and vulnerability (e.g., isolation from friends and family) onto themselves. Previous research has shown that postdisaster relocation and economic disadvantage can exacerbate the negative psychobiological consequences of the original event (Adams, Boscarino, & Galea, 2006; Cohen, Doyle, & Baum, 2006; McFarlane, 1990; Najarian, Goenjian, Pelcovitz, Mandel, & Najarian, 2001; Pérez-Sales, Cervellón, Vázquez, Vidales, & Gaborit, 2005). These findings suggest that displaced survivors of Hurricane Katrina may be at high risk for negative health consequences and prolonged difficulties in social and economic recovery due to the severity and uniqueness of their shared chronic (prolonged) and acute (daily) stress exposure (Osofsky, Osofsky, & Harris, 2007).

The Katrina sample was compared to a demographically matched comparison group that was not exposed to the hurricane. Comparisons were made between the Katrina and control groups on symptoms of depression, anxiety, psychological distress, aggression, and self-esteem, along with noninvasive measures (in saliva) of individual differences in HPA (cortisol) and SNS activity (sAA). We choose to focus on sex differences in biosocial stress responses, because males and females show baseline differences in certain types of problem behaviors, and may correspondingly develop unique behavioral and physiological patterns for responding to distress (Vigil, 2009). In general, females present greater symptoms of internalizing behaviors, such as depression and worrying, whereas males show greater externalizing behaviors, including higher levels of physical aggression and risk taking (Crick & Zahn-Waxler, 2003; Khan, Gardner, Prescott, & Kendler, 2002; Tolin & Foa, 2006; Vigil, 2008).

In laboratory studies, in comparison to women, men often show greater HPA and sympathetic activation in anticipation of and immediately following stressful stimuli, particularly events that involve a social-evaluative component (e.g., public speaking tasks; Kajantie & Philips, 2006; Kudielka et al., 1998; Uhart, Chong, Oswald, Lin, & Wand, 2006). Other research has found that cortisol spikes immediately following a traumatic experience (e.g., measurements taken at hospital admission centers) are more predictive of boys' rather than girls' subsequent risk for developing PTSD (Delahanty, Nugent, Christopher, & Walsh, 2005). More specific analyses suggest that women's stress systems are more sensitive to social exclusion, whereas men's stress systems are more sensitive to social-competitive situations (Salvador, 2005; Stroud, Salovey, & Epel, 2002). This research suggests that males may be at greater risk for adjustment problems following traumatic events that diminish the individual's relative social standing (Flinn, 2006; Sapolsky, 2004), such as significant material losses and lowering of social status that often result from natural disasters.

From an ultimate level of analysis or evolutionary perspective, normative patterns of behavioral distress, such as sadness and worrying behaviors have been described as *functional* for modifying the individual's behaviors (e.g., lowering risk-taking behavior in states of vulnerability) or the behaviors of others in ways that increase fitness (survival and reproductive prospects; Allen & Badcock, 2003; Fournier, Moskowitz, & Zuroff, 2002; Keller & Nesse, 2005; Sloman & Gilbert, 2000). For example, Vigil (2009) recently proposed a ''sociorelational'' framework that suggests that both sexes may respond to chronic stressors with an overall shift from the expression of dominant behaviors (e.g., assertiveness, aggression, extraversion) to submissive behaviors (e.g., depression and worrying behaviors) to signal trustworthiness to others and to elicit increased social support. Distress behaviors may operate in part to consolidate (strengthen) the reliability of one's relationships to gain additional needed support during adverse life experiences (see also Hagen, 2003).

In terms of sex differences, several theorists have suggested that modern-day masculine and feminine emotive styles (e.g., higher rates of crying and comforting behaviors in females and higher aggression and risk-taking behaviors in males) may by an artifact of unique social ecologies (network characteristics) and relationship demands in which ancestral males and females evolved (Geary, Byrd-Craven, Hoard, Vigil, & Numtee, 2003; Geary & Flinn, 2002; Taylor et al., 2000; Vigil, 2007, 2009). The basic distinction between male's and female's social dynamics pivots on the quality and quantity of relationships that members of each sex typically form and the level of investment behaviors that are needed to maintain different types of social networks. According to Vigil (2009), because girls and women tend to form fewer yet more intimate and more exclusive social relationships than males, on average, they may be similarly sensitive to display their vulnerability (e.g., via internalizing behaviors) in ways that advertise their trustworthiness and demonstrate their reciprocal investment to others (see also Geary et al., 2003; Vigil, 2007). Exaggerated displays of vulnerability and appeasement behaviors may in turn operate to strengthen the formation of more intimate and reliable, yet investment-intensive, social relationships among females.

In contrast, males may be more sensitive to *relax* expressions of trustworthiness cues in favor of behaviors that demonstrate the individual's physical abilities (e.g., strength and prowess) and other dominance-related traits to others. These types of displays are more observable (e.g., flashing body movements and displayed aggression) and require less investment (e.g., time and self-disclosure) to advertise to others, and may therefore be effective for forming and maintaining larger, less intimate, and less exclusive social relationships among coalitions of males (Vigil, 2009; see also Geary et al., 2003; Vigil, 2007). Because males and females show baseline differences in externalizing and internalizing behaviors (e.g., Crick & Zahn-Waxler, 2003; Khan et al., 2002; Vigil, 2008), we therefore predicted that each sex would manifest the expected shift from dominance to submissiveness behaviors differently; for instance, that displaced males would show lower levels of aggression and displaced females would show exaggerated symptoms of depression, compared to same-sex controls.

In all, we expected the main effects of exposure group on cortisol and sAA; specifically, the Katrina sample would show physical symptoms of chronic and acute stress exposure and hence lower salivary cortisol levels and higher sAA activity, respectively. We likewise examined the hypotheses that hurricane exposure moderated the relations between cortisol activity and psychological adjustment, as well as the earlier findings (e.g., El-Sheikh et al., 2008) that sAA activity may similarly moderate relations between cortisol and psychological adjustment. We were particularly interested in the potential moderating influence of these physiological measures on sex differences for anxiety, depression, and aggression (Gordis et al., 2008). We expected females living in the relocation camp would report the highest levels of depression when compared with nonexposed females and to males in general, and expected to find stronger correlations between sAA (a proxy for acute distress reactivity) and depression for females than for males. Finally, on the basis of the hypothesized pattern for people to respond to chronic stressors with an overall shift from dominant to submissiveness behaviors, we expected the Katrina sample would report lower aggression than would the controls and, given higher baseline aggression in males in general, that this difference would be most pronounced among displaced males.

Method

Event and Setting

Hurricane Katrina struck the North-Central Gulf coast (mostly Louisiana and Mississippi) on August 29, 2005. Although structural devastation and displacement was widespread, the most populated region affected by the hurricane was New Orleans, due in part to several breaches in the levees that protected the city from the surrounding Lake Pontchartrain (Travis, 2005). In addition to the direct loss of life, injury, and structural devastation caused by the hurricane, hundreds of thousands of people were either permanently or temporarily displaced from their communities. Many of these refugees were forced to seek shelter in government and community shelters, often being transported across multiple shelter sites throughout the United States. For the most severely affected individuals and families, government sponsored relocation camps were built to provide semipermanent (i.e., indefinitely provided) housing facilities. The largest of these camps, located near the city of Baton Rouge, Louisiana, was the site chosen to conduct the present study. At the time of our data collection, the camp, referred to by the residents and wider community as "Renaissance Village," consisted of approximately 500 travel trailers, each housing an average of three people (not including infants and young children). The site included a central tent and barracks area, which served as the local post office and cafeteria and where supplies (e.g., clothing) were dispersed; several small laundry houses; and a basketball court. Meals were served three times a day; residents typically retrieved these from the central building and brought them back to their trailers to eat with their families. Despite the frequent efforts taken to provide the residents with comfortable conditions (e.g., seasonal festivities), the general mood in the camp at the time of data collection was notably low.

Participants

A convenience sample of 62 adolescents (mean age = 14.9, *SD* = 2.0 years; 42 females) were recruited from the housing camp approximately 2 months after the hurricane and 3 weeks after the relocation camp opened, from October 28, 2005 through November 2, 2005. Researchers solicited participation by introducing themselves to the residents (door-to-door) and targeting families living in the trailer units with children. Parents were then asked if they wanted to participate in the study ''Child and Family Wellness after a Major Natural Disaster''; approximately 65% of families agreed to participate. Upon debriefing, the participants were given a small monetary payment. All the participants reported having been relocated from the New Orleans region. The control sample consisted of 53 adolescents (mean age = 15.3, *SD* = 2.1 years, 29 females) recruited from residential neighborhoods in mid-Missouri, USA and selected based on characteristics (e.g., age, race, socioeconomic status [SES]) matching the Katrina participants, from August, 2006 through January, 2007.

Prospective participants were informed about the project goals and their rights, as a research participant outlined by the University of Missouri Institutional Review Board, and once written consent was obtained from the children and their parents, participants were asked to complete a psychological survey. Parents of participants were also asked to provide demographic information. The latter confirmed the Katrina and control groups were similar in terms of race (both groups were over 93% African American) and SES; parents did not differ ($ps > .10$) for years of education (between 11 and 12 years), history of government assistance (e.g., welfare, food stamps; over 67%), or for indicators of wealth; the average yearly income of both samples was between $4,000 and $6,000, with current (posthurricane) financial assets of less than $200. The adolescent participants from the Katrina and control groups were similar ($ps > .15$) in age and proportion by sex.

Psychological Measures

The survey consisted of several standardized, self-report instruments designed to measure psychological well-being and family functioning in adolescents and young adults. For some instruments, entire scales were administered to allow for a direct comparison across normative samples; for other instruments, partial scales were used that consisted of items representing the instrument's subscales (this enabled greater breadth in the assessed traits). Each of the instruments have demonstrated reliable psychometric properties in young adult populations, and many of the scales have been used with similar adolescent samples who had recently experienced a large-scale disaster (e.g., Ginexi, Weihs, Simmens, & Hoy, 2000; Kreuger & Stretch, 2003; Sumer, Karanci, Berument, & Gunes, 2005; see also Joseph, 2000; Linton & Marriott, 1996).

Aggression was assessed by 8 items (2 items from each of the instrument's subscales) from the Aggression Questionnaire (Buss & Perry, 1992); the items were scored on a 5-point scale (overall $\alpha = .72$). Anxiety was assessed by the Revised Children's Manifest Anxiety Scale (Reynolds & Richmond, 1978), which consists of 37 items (each scored as yes or no responses; $\alpha = .87$). Depressive symptoms were assessed by the Center for Epidemiologic Studies Depression Scale (Radloff, 1977); the instrument consists of 20 items and each is scored on a 4-point scale ($\alpha = .86$). Subjective psychological distress was measured with the Impact of Events Scale–Revised (IES–R; Weiss & Marmar, 1997). The instrument consists of 22 items (which were revised to refer specifically to Hurricane Katrina; $\alpha = .93$) and four-factor subscales designed to measure symptoms of hyperarousal, disassociation, intrusive thoughts, and avoidance behaviors during the past week (scored on a 5-point scale). The Rosenberg Self-Esteem Scale (Rosenberg, 1965) was used to assess self-described global self-esteem; the instrument consists of 10 items and each is scored on a 4-point scale ($\alpha = .73$). Completion of the psychological measures and survey items usually took between 30 and 45 min.

Determination of Salivary Analytes

The two saliva samples were collected (immediately prior to and subsequent to completion of the survey measures) between the hours of 10:50 and 18:50 via a cotton swab version of the Salivette device (Sartstedt, Newton, NC); the samples were typically collected 60–90 min apart. Two saliva samples were collected, because of the impracticality of obtaining more collections under emergency field conditions and because they provided a more reliable means of measuring sAA and cortisol activity than single sample collections (for the implementation of similar techniques, see El-Sheikh et al., 2008). After collection, all samples were transported on ice and then stored frozen (−20 °C) at the University of Missouri until shipped overnight on dry ice to the Penn State University Behavioral Endocrinology Laboratory. They were stored frozen at −80 °C until assayed for cortisol and sAA. On the day of testing, all samples were centrifuged at 1500 g for 15 min to remove mucins.

Salivary alpha-amylase. Samples were assayed for sAA using a commercially available kinetic reaction assay (Salimetrics, State College, PA). The assay employs a chromagenic substrate, 2-chloro-*p*-nitrophenol, linked to maltotriose. The enzymatic action of sAA on this substrate yields 2-chloro-*p*-nitrophenol, which can be spectrophotometrically measured at 405 nm using a standard laboratory plate reader. The amount of sAA activity present in the sample is directly proportional to the increase (over a 2-min period) in absorbance at 405 nm. Results are computed in U/ml of sAA using the formula: [Absorbance difference per minute × total assay volume (328 ml) × dilution factor (200)]/[millimolar absorptivity of 2-chloro-*p*-nitrophenol (12.9) × sample volume (.008 ml) × light path (.97)]. Intra-assay variation (CV) computed for the mean of 30 replicate tests was less than 7.5%. Interassay variation computed for the mean of average duplicates for 16 separate runs was less than 6%. The correlation between the sAA measurements was modest ($r = .43$, $p < .01$), and correspondingly, the analyses used the average levels of the two sAA samples.

Cortisol. All samples were assayed for salivary cortisol using a highly sensitive enzyme immunoassay the U.S. Food and Drug Administration (510 k) cleared for use as an in vitro diagnostic measure of adrenal function (Salimetrics). The test used 25 μl of saliva (for singlet determinations), had a lower limit of sensitivity of .007 μg/dl, range of sensitivity from .007 to 1.8 μg/dl, and average intra- and interassay coefficients of variation of less than 5% and 10%, respectively. As expected, the correlations between the cortisol measurements were positive and strong ($r = .81$, $p < .01$) and thus the average of the two samples was used in all subsequent analyses.

Diurnal course of salivary cortisol and sAA. Both salivary cortisol and sAA show distinct and different diurnal profiles. Under normative-healthy conditions, HPA axis activity (on average) follows a diurnal pattern with cortisol levels high upon waking, then rising to a zenith within 30 min of waking, sharply declining by midday, and then gradually declining through the afternoon and evening (Kirschbaum & Hellhammer, 1989). By contrast, much less is known about diurnal variation of sAA. Recent and systematic studies with adults (ages 18–58 years) reveal sAA activity shows a pronounced decrease within 60 min after awakening and a steady increase of activity during the course of the day (Nater, Rohleder, Schlotz, Ehlert, & Kirschbaum, 2007). More specifically, Nater et al. (2007) report that sAA levels increased an average of 17% with each hour across the day. There was a quadratic curvature of the sAA change across the day, with sAA decelerating from its estimated zenith at 16:25. Noteworthy is the fact that, consistent with findings regarding acute stress reactivity, the pattern of production of sAA and cortisol across the day were uncorrelated. To the best of our knowledge, there are no studies of puberty-related differences in the sAA diurnal rhythm.

Data Analysis

Missing values for either of the two cortisol or sAA measures (3%) were replaced by the other; missing values for the psychological variables (7%) were replaced by the sample means for each sex. Preliminary analyses revealed that the participant's cortisol levels were positively related to his or her awakening time ($r = .23$, $p < .01$) and negatively related to collection time ($r = -.39$, $p < .01$). Cortisol was also positively related to participant's age ($r = .21$, $p < .05$), which may be a proxy for body mass. To help control for these factors, wake time, collection time, and age were entered as covariates (for analyses of covariance [ANCOVAs] and multivariate regressions) or were partialed out (for bivariate correlations) for all analyses involving cortisol. In contrast, wake time, collection time, and age were not related to sAA, and were thus not controlled for analyses using this analyte. Among the self-report measures, only psychological distress was related to wake time ($r = -.18$, $p = .06$) and collection time ($r = .19$, $p < .05$); these variables were therefore controlled for analyses that included symptoms of distress. The Katrina and control samples did not differ for wake time or saliva collection times ($ps > .23$); however, females reported a significantly later wake time than males, $t(113) = -3.79$, $p < .01$. Groups were compared using ANCOVAs; group differences are expressed as effect sizes (*d*s; mean difference/mean standard deviation). Multivariate correlations and descriptive plots were conducted using regression analyses.

Results

Group Differences in Psychological Functioning, Salivary Cortisol, and sAA

Table 1 shows the mean psychological and salivary scores for the Katrina and control groups separately for males and females, as well as the intervariable correlations for the entire sample. Two-way ANOVAs, using sex (males vs. females) and group (comparison vs. Katrina) as predictors for each of the psychological variables revealed significant Sex × Group interactions for depression, $F(1, 111) = 4.46$, $p < .05$, and aggression, $F(1, 111) = 6.15$, $p < .05$, and a trend for a significant interaction term for distress, $F(1, 109) = 3.51$, $p = .06$, but not for anxiety ($p = .21$) or self-esteem ($p = .47$). Follow-up regressions revealed that the interaction for depression emerged due to elevated depression scores for Katrina females compared to control females ($\beta = .32$, $p < .01$, $d = -0.68$), whereas no group difference for depression was found among males ($\beta = -.08$, $p = .61$, $d = 0.16$). In contrast, the interaction for aggression was due to lower aggression scores for Katrina males compared to control males ($\beta = -.55$, $p < .01$, $d = 1.31$), whereas no group difference for aggression was found among females ($\beta = -.13$, $p = .29$, $d = 0.26$). A main effect of sex was also found for self-esteem, $F(1, 111) = 4.58$, $p < .05$, $d = -0.36$, indicating lower self-esteem among males in both groups, compared to females. The Katrina group had significantly lower cortisol levels than the control group, $F(1, 108) = 5.74$, $p < .05$, $d = 0.41$; the main effect of sex and the Sample × Sex interaction term were not significant ($ps > .10$). Neither the interaction term nor the main effect terms for group or sex were significant for sAA ($ps > .19$).

Interactions Between sAA and Cortisol Levels and Psychological Functioning

Collapsing across sex, regression analyses were used to examine the relations between sAA and the psychological measures for the comparison and Katrina groups, and between cortisol and the psychological measures for both groups. For the sAA equations, each of the psychological measures was entered as the dependent variable, and the group, sAA, and Group × sAA interaction term were entered as independent variables. For the cortisol equations, group, cortisol, and the Group × Cortisol interaction term were entered as independent variables; wake time, collection time, and age were entered as covariates. The sAA analyses revealed a trend for a significant Sample × sAA interaction term for depression ($\beta = .33$, $p = .10$) and was due to a small negative relation between higher sAA and lower depressive symptoms for the control group ($\beta = -.29$, $p < .05$), but not for the Katrina

Table 1

Group Differences and Correlations for the Psychological and Salivary Analyte Variables

	Males		Females		Sample differences		Sex differences		Correlations							
	Control	Katrina	Control	Katrina	*t*	*d*	*t*	*d*	1	2	3	4	5	6	7	8
1. Sample	—	—	—	—	—	—	—	—								
2. Sex	—	—	—	—	—	—	—	—	—	—						
3. Depression	25.4 (10.0)	23.7 (11.6)	21.0 (9.2)	*27.6 (10.1)*	−1.75	−.33†	−0.16	−.03	.16†	.02	—					
4. Anxiety	22.8 (2.0)	21.2 (3.5)	21.1 (2.3)	20.9 (3.1)	1.73	.32†	2.00	.38*	−.16†	−.19*	−.45**	—				
5. Esteem	29.3 (5.3)	28.8 (6.8)	32.2 (3.5)	30.2 (5.0)	1.14	.22	−1.83	−.36†	−.11	.18*	−.49**	.36**	—			
6. Distress	61.1 (16.4)	54.6 (18.9)	52.0 (15.6)	*63.1 (19.6)*	−1.05	−.23	−0.85	−.02	.10	.08	.52**	−.38**	−.23*	—		
7. Aggression	97.5 (21.5)	*74.5 (12.6)*	95.9 (18.1)	91.0 (20.2)	3.01	.56**	−1.54	−.29	−.27**	.14	.27**	−.11	−.04	.17†	—	
8. Cortisol	0.20 (0.21)	0.15 (0.09)	0.20 (0.08)	*0.15 (0.11)*	2.42	.41*	1.34	.07	−.23*	−.13	.02	.11	.04	.13	.17†	—
9. Amylase	54.6 (51.7)	57.3 (43.4)	42.8 (28.5)	*66.1 (48.1)*	−1.84	−.35†	−0.08	−.02	.17†	.01	−.07	−.08	−.02	−.01	.01	−.02

Note. Values are mean scores; values in parentheses are standard deviations. Italicized values indicate significant ($p < .05$) sample differences across males or females. Group differences for distress were computed by entering collection time and wake time as covariates; group differences for cortisol were computed by entering collection time, wake time, and age as covariates. Partial correlations were computed for distress (controlling for collection time and wake time) and for cortisol (controlling for collection time, wake time, and age).

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

group (β = .02, $p = .88$). The cortisol analyses revealed significant Group × Cortisol interaction terms for self-esteem (β = .47, $p > .01$), anxiety (β = .33, $p > .05$), and distress (β = −.34, $p > .05$), and a trend for an interaction for depression (β = −.31, $p = .06$). Follow-up regressions revealed that the interactions for self-esteem and anxiety were due to small positive relations between cortisol and self-esteem (β = .36, $p < .05$) and between cortisol and anxiety (β = .26, $p = .09$) for the Katrina group, and a small negative relation between cortisol and self-esteem (β = −.26, $p = .09$) for the control group; the relation between cortisol and anxiety was not significant for the control group ($p = .64$). In contrast, the interactions for distress and depression were due to small positive relations between cortisol and distress (β = .39, $p < .01$) and between cortisol and depression (β = .26, $p = .09$) for the control group, and nonsignificant relations for the Katrina group (*p*s = .78 and .21, respectively).

Collapsing across groups and sex, regressions were then used to examine the potential interactions between cortisol and the psychological measures for individuals with varying sAA activity levels. For these equations, each of the psychological measures was entered as the dependent variable, and cortisol, sAA, and the Cortisol × sAA interaction term were entered as independent variables; wake time, collection time, and age were entered as covariates. These analyses revealed a significant Cortisol × sAA interaction term for self-esteem (β = .43, $p < .05$) and a trend for an interaction for symptoms of anxiety (β = .33, $p = .10$); the interaction terms were not significant for depression ($p = .73$), distress ($p = .84$), or aggression ($p = .88$). Follow-up regressions revealed that the significant interaction emerged as a result a moderate positive relation between cortisol and self-esteem for individuals with sAA levels above the overall mean (β = .43, $p < .05$) but not for those with sAA levels below the mean (β = −.04, $p = .74$).

Sex Differences in the Relations Between Psychological Functioning and Cortisol and sAA Levels

Regressions were then used to examine the relations between each of the two analytes and the psychological scores, separately among males and females. For these equations, collapsing across groups, each of the psychological measures was entered as the dependent variable, and sex, the salivary analyte (i.e., cortisol or sAA), and the Sex × Analyte interaction term were entered as independent variables; wake time, collection time, and age were entered as covariates for the cortisol equations. These analyses revealed a trend for a Cortisol × Sex interaction for psychological distress, β = −.51, $p = .10$, and was due to a significant relation between higher cortisol and increased distress among males (β = .36, $p < .05$) and a nonsignificant relation among females ($p = .99$). None of the remaining main effects for cortisol, or the Cortisol × Sex interaction terms were significant. In contrast, regressions revealed a significant sAA × Sex interaction term for depression (β = 1.18, $p < .01$), and a trend for a sAA × Sex interaction for psychological distress (β = .64, $p = .06$). As shown in Figure 1, the significant interaction was due to a negative relation between sAA and depression in males (β = −.47, $p < .01$), and a trend for a positive relation between sAA and depression in females (β = .21, $p = .08$). No other effects of sAA were significant for the remaining psychological variables.

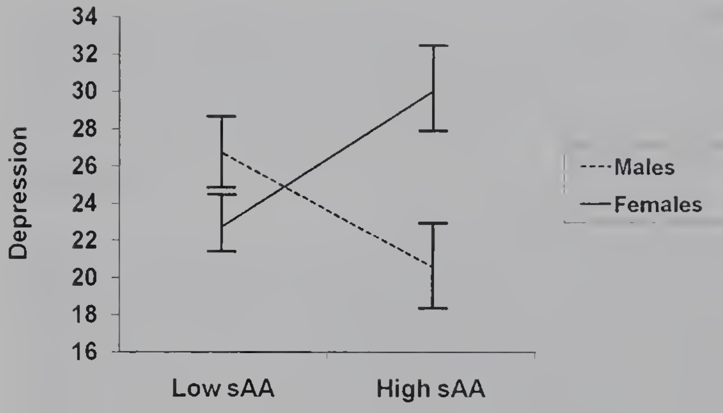


Figure 1. Plot of significant Sex × Salivary Alpha Amylase (sAA) interaction for depression. Slopes represent unstandardized regression coefficients.

Discussion

This study examined psychological functioning and concurrent cortisol and sAA activity levels in adolescents who experienced pervasive stressors resulting from significant social and material losses during and following Hurricane Katrina. In comparison to controls, both sexes in the Katrina sample showed moderately lower cortisol levels and slightly higher sAA activity; however, these differences were more pronounced among displaced females. The finding of lower cortisol among the displaced youth is consistent with other studies showing that chronic stress exposure is associated with blunted cortisol activity, potentially reflecting a general habituation or desensitization of this stress response (Flinn et al., 1996; Miller et al., 2007). Higher sAA activity (reflecting more acute stress reactivity) among the Katrina youth may reflect group differences in daily stressors, such as the temporary housing facilities and the lack of security that were the basis of common complaints from the Katrina youth.

Our findings revealed that hurricane exposure moderated the relations between cortisol activity and psychological adjustment; however, these effects were small and variable and should therefore be considered preliminary findings. Future research should benefit from more detailed theoretical models that differentiate the influences of HPA activity on different aspects of internalizing behaviors, including the respective roles on self-esteem, depression, and symptoms of anxiety, among individuals who have experienced differing levels of stress exposure. We also found support for the hypothesis that sAA moderates the relation between cortisol and internalizing behaviors; lower cortisol was associated with lower self-esteem, but only among individuals with above average sAA reactivity. This finding complements recent laboratory studies (e.g., Fortunato et al., 2008), and suggests that individual differences in HPA axis activity and psychological functioning may be associated with differential sensitivity or functioning of the SNS. Overall, however, the results are most compatible with a modified version of Bauer et al. (2002) interactionist model (p. 106), and suggest that individuals with high sAA activity and low cortisol may be at greater risk for mood problems, whereas high sAA combined with high cortisol may be associated with psychological resiliency.

In terms of sex differences, and consistent with research among normative (nontraumatized) adolescents (e.g., Ge et al., 2003; Rushton, Forcier, & Schectman, 2002; see also Vigil, 2008), the females in our study reported higher symptoms of depression than did the males. Individuals of both sexes and in both groups reported higher levels of depressive symptoms than are found in normative samples (e.g., see Brooks & Kutcherm, 2001, p. 365), but these symptoms were particularly high among the Katrina females. The pattern is consistent with a sociorelational perspective of distress behaviors, and the evolutionary hypothesis that people in general, and especially females, may respond to negative life events with behaviors that exaggerate the displays of vulnerability and increase the perception of trustworthiness to other people (Vigil, 2009). From this perspective, certain types of behavioral distress (e.g., worrying, crying, feelings of hopelessness) may be functional for consolidating one's social network and evoking increased emotional support from intimate confidantes (e.g., family members and close friendships), in coordination with adverse life experiences. These are the types of events that demand the greatest social support from others, and behaviors that explicate individual vulnerability may be particularly effective for evoking dependable and continuous forms of investment from others (Vigil, 2009; see also Hagen, 2003).

Females may be particularly sensitive to display cues of trustworthiness (e.g., via sadness behaviors) and to emotionally invest in their relationships (e.g., via comforting behaviors), compared to males, because these behaviors were beneficial for maintaining more intimate and exclusive dyadic relationships (e.g., see Geary et al., 2003; Vigil, 2007, 2009; see also Geary & Flinn, 2002; Taylor et al., 2000). Females may be particularly vulnerable during adolescence and reproductive maturation and especially in situations of unreliable social support; these are conditions that can result in sexual exploitation and pregnancy. Adolescence is a time when sex differences in the display of trustworthiness cues (e.g., via sadness behaviors) are predicted to emerge, and to function to maintain social-support systems (see Del Giudice, 2009; Geary et al., 2003; Vigil, 2009). This hypothesis is consistent with research on depression during adolescence that shows that symptoms tend to increase in females, but not males, between 12 and 13 years in U.S. samples (e.g., Ge, Conger, & Elder, 2001; Ge et al., 2003). Likewise, recent research shows a relation between HPA activity and intimate relationship dynamics (e.g., quality of relationships) in adolescent females, but not females in middle childhood, or males of either age group (Booth, Granger, & Shirtcliff, 2008).

Reported levels of aggression were similarly higher than that found in normative samples for all groups (Archer, Kilpatrick, & Bramwell, 1995, p. 376), except the Katrina males. Aggression for these males was in the normative range but substantially lower than that reported by the control males or by females of either group. Although counterintuitive, this finding was predicted based on higher normative (no-trauma) levels of aggressive behaviors in males in difficult circumstances and the hypothesis that people respond to chronic, noninterpersonal victimization, stressors with decreased dominance and increased submissive behaviors (Allen & Badcock, 2003; Fournier et al., 2002; Sloman & Gilbert, 2000; Vigil, 2009). Similar patterns emerge in male primates under chronic stress and following loss of status (Sapolsky, 2004, 2005). From a sociorelational perspective, aggressive behaviors may be similarly interpreted as a form of social behavior that may be more adaptive for males than for females under normal (no-trauma) conditions but are associated with added costs when males experience events that decrease their social status or the ability to protect themselves (Vigil, 2009).

Another possibility for our finding of relatively lower aggression in affected males may be that the males in this study are unrepresentative of more general samples; males willing to accept government assistance may be less aggressive than displaced males who are not living in a relocation camp. Lower aggression among displaced males may also reflect environmental rather than conditional or individual differences. For example, the Katrina camp environment may be more mundane than the environment in which the Katrina males were situated prior to the hurricane or in comparison to the acute stressors (e.g., overcrowding conditions at the Superdome, temporary disaster facility), immediately following the hurricane. In any case, males are more sensitive to changes in social standing than are females (Geary, 1998), which may explain our finding of the stronger concordance between cortisol and symptoms of psychological distress in males than among females (for similar results, see Delahanty et al., 2005).

Finally, the results support the hypothesis that sex moderates the relation between sAA and symptoms of depression, such that higher sAA among the youth in both groups was related to lower symptoms of depression (e.g., greater confidence and reported elation) in males and slightly higher depressive symptoms among females. The findings are consistent with a recent study that showed that sAA is distinctly associated with positive affect and overall approach behaviors (Fortunato et al., 2008), and research indicating normative sex differences in behavioral reactivity (i.e., to respond to mild stressors with more externalizing vs. internalizing behaviors; Vigil, 2008). Moreover, our findings suggest that sAA may be a promising marker for identifying depressive symptoms in adolescents, but they are qualified by the collection of only two saliva samples. It is possible, for instance, that sAA activity may follow a diurnal rhythm (see findings by Nater et al., 2006) similar to patterns observed in cortisol. Still, we found no evidence for such a pattern of sAA in our study using controls for the independent influence of collection time and awakening time. Systematic studies are needed to examine the effectiveness of sAA for measuring symptoms of depression among children exposed to ordinary and extraordinary stressors.

Limitations

Future research on individual and group differences in stress responses following large-scale disasters will benefit from addressing specific constraints in the design of these studies. For example, although the window of time we used to collect the saliva samples was reasonable under the constraints of our field study, the use of practical saliva collection techniques that enable researchers to track hormonal fluctuations in the morning and across the day under emergency field conditions is needed. This research is particularly complicated when working with transient families and individuals with low education, as was the case in the current study. It was unfortunate, although understandable, that pre-trauma information was not available, particularly for baseline physical or mental health, and non-trauma-related stressors in either the Katrina or the control group. It would have also been helpful to have been able to examine the role of pubertal timing, especially among females, for examining more specific hypotheses from the theoretical models. The current study was also limited by the small sample size and homogeneous characteristics (e.g., low SES) of our samples. Future research will benefit from assessments of psychological functioning of males and females from representative samples that include adolescents with diverse family and socioeconomic backgrounds. Additional limitations were the single source (self-report) of data collection, lack of information on current illnesses and medication use, and the relatively wide age range of participants that were assessed. It is certainly possible that children undergo systematic psychological and physiological changes that may operate in conjunction with distinct psychoneuroendocrine and behavioral patterns of stress response, at different stages of development (e.g., see Vigil, 2009).

Despite these limitations, the study contributes to our understanding of normal psychobiological distress in males and females, and it is the first study to demonstrate sex differences in psychological functioning and salivary cortisol and sAA activity among adolescents that have experienced pervasive stressors that included significant disruptions in social relationships and widespread loss of property caused by a major natural disaster. It is unclear at this time, however, how these types of experiences and stress recovery mechanisms may actually affect normative and long-term development. Recent laboratory studies have shown that, in comparison to younger children, adolescents may be particularly sensitive to social evaluation and stressors that reduce the individual's social standing (Booth et al., 2008; Stroud et al., 2009). Lower sociometric competitiveness caused by significant social and material losses may in turn have an exacerbating effect for adolescents already in low-resource environments. Based on our findings, such disruptions may be manifested through divergent behavioral patterns in males and females, which in theory may be associated with remedial social benefits, as well as potential mental and physical health-related consequences.

References

Adams, R. E., Boscarino, J. A., & Galea, S. (2006). Social and psychological resources and health outcomes after the World Trade Center disaster. *Social Science Medicine*, *62*, 176–188.

Allen, N. B., & Badcock, P. B. T. (2003). The social risk hypothesis of depressed mood: Evolutionary, psychosocial, and neurobiological perspectives. *Psychological Bulletin*, *129*, 887–913.

Archer, J., Kilpatrick, G., & Bramwell, R. (1995). Comparison of two aggression inventories. *Aggressive Behavior*, *21*, 371–380.

Bauer, A. M., Quas, J. A., & Boyce, W. T. (2002). Associations between physiological reactivity and children's behavior: Advantages of a multisystem approach. *Journal of Developmental & Behavioral Pediatrics*, *23*, 102–113.

Booth, A., Granger, D. A., & Shirtcliff, E. A. (2008). Gender- and age-related differences in the association between social relationship quality and trait levels of salivary cortisol. *Journal of Research on Adolescence*, *18*, 239–260.

Brooks, S. J., & Kutcherm, S. (2001). Diagnosis and measurement of adolescent depression: A review of commonly utilized instruments. *Journal of Child and Adolescent Psychopharmacology*, *11*, 341–376.

Burke, H. M., Fernald, L. C., Gertler, P. J., & Adler, N. E. (2005). Depressive symptoms are associated with blunted cortisol stress responses in very low-income women. *Psychosomatic Medicine*, *67*, 211–216.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*, 452–459.

Chae, E., Kim, T. W., Rhee, S., & Henderson, T. D. (2005). The impact of flooding on the mental health of affected people in South Korea. *Community Mental Health Journal*, *41*, 633–645.

Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. *Journal of the American Medical Association*, *267*, 1244–1252.

Cohen, S., Doyle, W. J., & Baum, A. (2006). Socioeconomic status is associated with stress hormones. *Psychosomatic Medicine*, *68*, 414–420.

Cohen, S., Frank, E., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Jack M. Gwaltney, J. M. (1998). Types of stressors that increase susceptibility to the common cold in healthy adults. *Health Psychology*, *17*, 214–223.

Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: Current progress and future challenges. *Developmental Psychopathology*, *15*, 719–742.

Delahanty, D. L., Nugent, N. R., Christopher, N. C., & Walsh, M. (2005). Initial urinary epinephrine and cortisol levels predict acute PTSD symptoms in child trauma victims. *Psychoneuroendocrinology*, *30*, 121–128.

Del Giudice, M. (2009). Sex, attachment, and the development of reproductive strategies. *Behavioral and Brain Sciences*, *32*, 1–67.

Ehlert, U., Gaab, J., & Heinrichs, M. (2001). Psychoneuroendocrinological contributions to the etiology of depression, posttraumatic stress disorder, and stress-related bodily disorders: The role of the hypothalamus-pituitary-adrenal axis. *Biological Psychology*, *57*, 141–152.

Ellis, B. J., Jackson, J. J., & Boyce, W. T. (2006). The stress response systems: Universality and adaptive individual differences. *Developmental Review*, *26*, 175–212.

El-Sheikh, M., Erath, S. A., Buckhalt, J. A., Granger, D. A., & Mize, J. (2008). Cortisol and children's adjustment: The moderating role of sympathetic nervous system activity. *Journal of Abnormal Child Psychology*, *36*, 601–611.

Flinn, M. V. (2006). Evolution and ontogeny of stress response to social challenges in the human child. *Developmental Review*, *26*, 138–174.

Flinn, M. V., & England, B. G. (2003). Childhood stress: Endocrine and immune responses to psychosocial events. In J. M. Wilce (Ed.), *Social and cultural lives of immune systems* (pp. 107–147). London: Routledge.

Flinn, M. V., Quinlan, R. J., Turner, M. T., Decker, S. D., & England, B. G. (1996). Male-female differences in effects of parental absence on glucocorticoid stress response. *Human Nature*, *7*, 125–162.

Fortunato, C. K., Dribin, A. E., Granger, D. A., & Buss, K. A. (2008). Salivary alpha- amylase and cortisol in toddlers: Differential relations to affective behavior. *Developmental Psychobiology*, *50*, 807–818.

Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2002). Social rank strategies in hierarchical relationships. *Journal of Personality and Social Psychology*, *83*, 425–433.

Ge, X., Conger, R. D., & Elder, G. H. (2001). Pubertal transition, stressful life events, and the emergence of gender differences in adolescent depressive symptoms. *Developmental Psychology*, *37*, 404–417.

Ge, X., Kim, I. J., Brody, G. H., Conger, R. D., Simons, R. L., Gibbons, F. X., et al. (2003). It's about timing and change: Pubertal transition effects on symptoms of major depression among African American youths. *Developmental Psychology, 39*, 430–439.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences.* Washington, DC: American Psychological Association.

Geary, D. C., Byrd-Craven, J., Hoard, M. K., Vigil, J., & Numtee, C. (2003). Evolution and development of boys' social behavior. *Developmental Review, 23*, 444–470.

Geary, D. C., & Flinn, M. V. (2002). Sex differences in behavioral and hormonal response to social threat: Commentary on Taylor et al. (2000). *Psychological Review, 109*, 745–750.

Ginexi, E. M., Weihs, K., Simmens, S. J., & Hoy, D. R. (2000). Natural disaster and depression: A prospective investigation of reactions to the 1993 Midwest floods. *American Journal of Community Psychology, 28*, 495–518.

Gordis, E. B., Granger, D. A., Susman, E. J., & Trickett, P. K. (2006). Asymmetry between salivary cortisol and alpha-amylase reactivity to stress: Relation to aggressive behavior in adolescents. *Psychoneuroendocrinology, 31*, 976–987.

Gordis, E. B., Granger, D. A., Susman, E. J., & Trickett, P. K. (2008). Salivary alpha amylase- cortisol asymmetry in maltreated youth. *Hormones and Behavior, 53*, 96–103.

Granger, D. A., Kivlighan, K. T., El-Sheikh, M., Gordis, E., & Stroud, L. R. (2007). Salivary alpha-amylase in biobehavioral research: Recent developments and applications. *Annals of the New York Academy of Sciences, 1098*, 122–144.

Granger, D. A., Weisz, J. R., McCracken, J. T., Ikeda, S., & Douglas, P. (1996). Reciprocal influences among adrenocortical activation, psychosocial processes, and clinic-referred children's short-term behavioral adjustment. *Child Development, 67*, 3250–3262.

Hagen, E. H. (2003). The bargaining model of depression. In P. Hammerstein (Ed.), *The genetic and cultural evolution of cooperation* (pp. 95–123). Cambridge, MA: MIT Press.

Joseph, S. (2000). Psychometric evaluation of Horowitz's Impact of Event Scale: A review. *Journal of Traumatic Stress, 13*, 101–113.

Kagan, J., Snidman, N., Arcus, D., & Reznick, J. S. (1994). *Galen's prophecy: temperament in human nature.* New York: Basic Books.

Kajantie, E., & Philips, D. I. W. (2006). The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology, 31*, 151–178.

Keller, M. C., & Nesse, R. M. (2005). Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. *Journal of Affective Disorders, 86*, 27–35.

Khan, A. A., Gardner, C. O., Prescott, C. A., & Kendler, K. S. (2002). Gender differences in the symptoms of major depression in opposite-sex dizygotic twin pairs. *American Journal of Psychiatry, 159*, 1427–1429.

Kirschbaum, C., & Hellhammer, D. K. (1989). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology, 19*, 313–333.

Klimes-Dougan, B., Hastings, P. D., Granger, D. A., Usher, B., & Zahn-Waxler, C. (2001). Adrenocortical activity in at risk and normally developing adolescents: Individual differences in salivary cortisol response to social challenges, basal levels, and circadian rhythms. *Development and Psychopathology, 13*, 695–720.

Kreuger, L., & Stretch, J. (2003). Identifying and helping long-term child and adolescent disaster victims: Model and methods. *Journal of Social Service Research, 30*, 93–108.

Kudielka, B. M., Hellhammer, J., Hellhammer, D. H., Wolf, O. T., Pirke, K., Varadi, E., et al. (1998). Sex differences in endocrine and psychological responses to psychosocial stress in healthy elderly subjects and the impact of a 2-week dehydroepiandrosterone treatment. *Journal of Clinical Endocrinology and Metabolism, 83*, 1756–1761.

Kunz-Ebrecht, S. R., Mohamed-Ali, V., Feldman, P. J., Kirschbaum, C., & Steptoe, A. (2003). Cortisol responses to mild psychological stress are inversely associated with proinflammatory cytokines. *Brain, Behavior, and Immunity, 17*, 373–383.

Linton, K. E., & Marriott, R. G. (1996). Self-esteem in adolescents: Validation of the state self-esteem scale. *Personality and Individual Differences, 21*, 85–90.

McFarlane, A. C. (1990). An Australian disaster: The 1983 bushfires. *International Journal of Mental Health, 19*, 36–47.

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-andrenocortical axis in humans. *Psychological Bulletin, 133*, 25–45.

Misra, M., Miller, K. K., Almazan, C., Ramaswamy, K., Lapcharoensap, W., Worley, M., et al. (2004). Alterations in cortisol secretory dynamics in adolescent girls with anorexia nervosa and effects on bone metabolism. *Journal of Clinical Endocrinology and Metabolism, 89*, 4972–4980.

Najarian, L. M., Goenjian, A. K., Pelcovitz, D., Mandel, F., & Najarian, B. (2001). The effect of relocation after a natural disaster. *Journal of Traumatic Stress, 14*, 511–526.

Nater, U. M., La Marca, R., Florin, L., Moses, A., Langhans, W., Koller, M. M., et al. (2006). Stress-induced changes in salivary alpha-amylase activity—associations with adrenergic activity. *Psychoneuroendocrinology, 31*, 49–58.

Nater, U. M., Rohleder, N., Schlotz, W., Ehlert, U., & Kirschbaum, C. (2007). Determinants of the diurnal course of salivary alpha-amylase. *Psychoneuroendocrinology, 32*, 392–401.

Osofsky, J. D., Osofsky, H. J., & Harris, W. W. (2007). Katrina's children: Social policy considerations for children in disasters. *Social Policy Report, 21*, 3–18.

Pérez-Sales, P., Cervellón, P., Vázquez, C., Vidales, D., & Gaborit, M. (2005). Post-traumatic factors and resilience: The role of shelter management and survivors' attitudes after the earthquakes in El Salvador (2001). *Journal of Community & Applied Social Psychology, 15*, 368–382.

Radloff, L. S. (1977). The CES–D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Reynolds, C. R., & Richmond, B. O. (1978). ''What I think and feel'': A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6*, 271–280.

Rohleder, N., Nater, U. M., Wolf, J. M., Ehlert, U., & Kirschbaum, C. (2004). Psychosocial stress induced activation of salivary alpha-amylase. An indicator of sympathetic activity? *Annals of the New York Academy of Sciences, 1032*, 258–263.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rotton, J., & Dubitsky, S. S. (2002). Immune function and affective states following a natural disaster. *Psychological Reports, 90*, 521–524.

Rushton, J. L., Forcier, M., & Schectman, R. M. (2002). Epidemiology of depressive symptoms in the National Longitudinal Study of Adolescent Health. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 199–205.

Salvador, A. (2005). Coping with competitive situations in humans. *Neuroscience and Biobehavioral Reviews, 29*, 195–205.

Sapolsky, R. M. (2004). Social status and health in humans and other animals. *Annual Review of Anthropology, 33*, 393–418.

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science, 308*, 648–652.

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin, 130*, 601–630.

Sloman, L., & Gilbert, P. (2000). *Subordination and defeat. An evolutionary approach to mood disorders.* Mahwah, NJ: Erlbaum.

Stroud, L. R., Foster, E., Papandonatos, G., Handwerger, K., Granger, D. A., Kivlighan, K. T., et al. (2009). Stress response and the adolescent transition: Performance versus peer social rejection stress. *Development and Psychopathology, 21*, 47–68.

Stroud, L. R., Salovey, P., & Epel, E. S. (2002). Sex differences in stress responses: Social rejection versus achievement stress. *Biological Psychiatry, 52*, 318–327.

Sumer, N., Karanci, A. N., Berument, S. K., & Gunes, H. (2005). Personal resources, coping self-efficacy, and quake exposure as predictors of psychological distress following the 1999 earthquake in Turkey. *Journal of Traumatic Stress, 18*, 331–342.

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review, 107*, 411–429.

Tolin, D. F., & Foa, E. B. (2006). Sex Differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin, 132*, 959–992.

Travis, J. (2005). Hurricane Katrina: Scientists' fears come true as hurricane floods New Orleans. *Science, 309*, 1656–1659.

Uchino, B. N., Cacioppo, J. T., & Kiecolt-Glaser, J. K. (1996). The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin, 119*, 488–531.

Uhart, M., Chong, R. Y., Oswald, L., Lin, P., & Wand, G. S. (2006). Gender differences in hypothalamic-pituitary-adrenal (HPA) axis reactivity. *Psychoneuroendocrinology, 31*, 642–652.

Vigil, J. M. (2007). Asymmetries in the friendship preferences and social styles of men and women. *Human Nature, 18*, 143–161.

Vigil, J. M. (2008). Sex differences in affect behaviors, desired social responses, and accuracy at understanding the social desires of other people. *Evolutionary Psychology, 6*, 506–522.

Vigil, J. M. (2009). A socio-relational framework of sex differences in the expression of emotion. *Behavioral and Brain Sciences, 32*, 375–390.

Weinstein, N. D., Lyon, J. E., Rothman, A. J., & Cuite, C. A. (2000). Changes in perceived vulnerability following a natural disaster. *Journal of Social and Clinical Psychology, 19*, 372–395.

Weiss, D., & Marmar, C. (1997). The Impact of Event Scale–Revised. In J. Wilson & T. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 399–411). New York: Guilford.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1241–1259

Children of Katrina: Lessons Learned About Postdisaster Symptoms and Recovery Patterns

Mindy E. Kronenberg, Tonya Cross Hansel, Adrianne M. Brennan, Howard J. Osofsky, and Joy D. Osofsky
Louisiana State University Health Sciences Center

Beverly Lawrason
St. Bernard Parish Public Schools

Trauma symptoms, recovery patterns, and life stressors of children between the ages of 9 and 18 (n = 387) following Hurricane Katrina were assessed using an adapted version of the National Child Traumatic Stress Network Hurricane Assessment and Referral Tool for Children and Adolescents (National Child Traumatic Stress Network, 2005). Based on assessments 2 and 3 years after the hurricane, most children showed a decrease in posttraumatic stress and depression symptoms over time. Students were also classified into outcome trajectories of stress resistant, normal response and recovery, delayed breakdown, and breakdown without recovery (A. S. Masten & J. Obradovic, 2008). Age, gender, and life stressors were related to these recovery patterns. Overall, the findings highlight the importance of building and maintaining supportive relationships following disasters.

> *When we evacuated, I was glued to the TV watching the destruction. I asked my parents for help, but they were as bewildered as I was. I realized I had to grow up fast to be able to help myself and also support them. My father still does not have his job back. I don't know what the future holds, but after surviving Katrina, I am sure I can handle it.* (14-year-old high-school student 1 year after Hurricane Katrina)

Immediately after Hurricane Katrina struck the Gulf Coast on August 29, 2005, both physical and psychological recovery began. Hurricane Katrina has been described as one of the worst natural disasters in U.S. history (Knabb, Rhome, & Brown, 2006), and, 4 years later, it continues to impact residents and communities in Louisiana. For example, St. Bernard Parish, an adjacent parish (county) to New Orleans in which this study was conducted, was completely devastated by the hurricane with the entire region damaged by flooding and a large oil spill. At the time of this study, recovery in St. Bernard Parish had been slow with approximately 50% of the population returning to the parish (St. Bernard Parish Government, 2008) and 47% of the children returning to the schools (St. Bernard Parish Public Schools, 2008).

Disasters with a slow recovery, such as Hurricane Katrina, can result in acute and chronic psychological effects (Kessler, Galea, Jones, & Parker, 2006; H. J. Osofsky, Osofsky, Kronenberg, Brennan, & Hansel, 2009; Weems et al., 2007), which may negatively impact the child's normal developmental trajectory (Pynoos, Steinberg, & Piacentini, 1999; Shaw, 2000). Of particular interest to child development is the occurrence of large-scale traumas, such as Hurricane Katrina, which affect not only the individual but also the multiple systems, including microsystems and exosystems, in which children develop (Bronfenbrenner, 1986; Masten & Obradovic, 2008). For many children of Katrina, their once-thriving neighborhoods, grocery stores, and playgrounds were no longer functional; many children experienced multiple moves and changes in schools as well as parental unemployment (Osofsky, Osofsky, & Harris, 2007; H. J. Osofsky et al., 2009). The outcomes for these children are still under question.

This research was supported by grants from the Robert Wood Johnson Foundation and Baptist Community Ministries. Much appreciation is expressed to the trauma team in the Department of Psychiatry at Louisiana State University Health Sciences Center and to St. Bernard Parish Public Schools. We would like to acknowledge Michelle Bacon Moore and Erika L. Bocknek for their assistance with references and editing.

Correspondence concerning this article should be addressed to Joy D. Osofsky, Departments of Pediatrics and Psychiatry, Louisiana State University Health Sciences Center, 1542 Tulane Avenue, 2nd floor, New Orleans, LA 70112. Electronic mail may be sent to josofs@lsuhsc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0015

Traditionally, child-focused disaster research examines postdisaster symptomatology. For example, the majority of the current literature suggests that children who experience hurricanes are at a higher risk for symptoms of depression, anxiety, and posttraumatic stress disorder (PTSD; Goenjian et al., 2001; Kessler et al., 2006; J. D. Osofsky, Osofsky, Kronenberg, & Cross, 2009; H. J. Osofsky et al., 2007); however, recent theoretical research has focused on examining patterns of recovery (Bonanno, 2004; Bonanno & Mancini, 2008; Masten & Obradovic, 2008). The problem with placing emphasis only on symptoms following disasters is that the focus is on psychopathology that limits attention to the full array of recovery. By examining recovery patterns, researchers can turn their attention to the multiple outcomes that follow disasters including posttraumatic growth and continued normal development, as well as maladaptation.

This study approaches disaster research by examining not only the symptomatology endorsed by children, who have experienced Hurricane Katrina but also the recovery patterns of these children and life stressors that influence these patterns. More specifically, this study examined longer term recovery and focused on the responses provided by children in the 2nd and 3rd years following Hurricane Katrina. Variable-focused analyses examined the course of trauma-related symptoms of depressed mood and posttraumatic stress over time, whereas person-focused analyses compared groups of students in 4th–12th grades who fit theoretical patterns of recovery described in the literature as either positive or negative outcome trajectories (Masten & Obradovic, 2008). The effects of life stressors, age, and gender on the course of post-Katrina trauma symptoms and recovery patterns were also examined.

Impact of Natural Disasters on School-Age Children and Adolescents (PTSD and Depressive Symptoms)

A number of disaster studies have documented symptoms of PTSD in children who have experienced natural disasters including earthquakes, tsunamis, and hurricanes (Goenjian et al., 2005; John, Russell, & Russell, 2007; Kolaitis et al., 2003; La Greca, Silverman, Vernberg, & Prinstein, 1996; Lonigan, Shannon, Taylor, Finch, & Sallee, 1994; Piyasil et al., 2007; Pynoos et al., 1993). Many of these studies have reported comorbid symptoms of depression in children following natural disasters. Although there is overlap in symptomatology of both PTSD and depression, including anhedonia, sleep difficulties, problems with concentration, irritability, and a restricted range of affect (American Psychiatric Association, 2000), the research has been clear in demonstrating the distinct presence of each disorder following disasters (Goenjian et al., 2001; Kolaitis et al., 2003; Roussos et al., 2005). For example, a study exploring PTSD and depression in children between the ages of 7 and 17 who experienced a super-cyclone in India found that although PTSD and depression were significantly correlated, most children with PTSD did not meet criteria for depression, and 55.7% of children with a diagnosis of depression did not meet criteria for PTSD (Kar et al., 2007).

The relation between PTSD symptoms and disaster-specific aspects of trauma has been well documented in the literature. For example, a study of 16- and 17-year-old children following a 1999 earthquake in Greece (Roussos et al., 2005) as well as studies of 13-year-olds following Hurricane Mitch in Nicaragua (Goenjian et al., 2001, 2005) showed that exposure to a natural disaster was consistently related to increases in the severity of PTSD symptoms. In a study of over 5,000 children aged 9–19 who experienced displacement and damage to their homes as a result of Hurricane Hugo, PTSD was associated with traumatic experiences (Lonigan et al., 1994). Similarly, Russoniello et al. (2002) found that 9- to 12-year-old children whose homes were flooded as a result of Hurricane Floyd were three times more likely to have symptoms of PTSD compared with those whose homes did not flood. Hamada, Kameoka, Yanagida, and Chemtob (2003) reported that 6- to 12-year-old children who experienced Hurricane Iniki were more likely to report posttraumatic symptoms if they felt that their lives or the lives of others were threatened at the time of the hurricane.

Neuner, Schauer, Catani, Ruf, and Elbert (2006) assessed 64 tsunami survivors in Sri Lanka, aged 8–14, and found that previous traumas, including exposure to war, domestic violence, community violence, medical treatment, physical abuse, and natural disaster, were associated with increased posttraumatic stress symptoms. Similarly, Garrison, Weinrich, Hardin, Weinrich, and Wang (1993) assessed 1,264 children, aged 11–17, following Hurricane Hugo and found that experiencing previous violent, traumatic events was associated with increased likelihood of PTSD.

Unlike symptoms of PTSD, depression has not been found to be consistently related to the level of disaster exposure or proximity. Depression has been associated with several different factors including reported difficulties at home following the disaster (Roussos et al., 2005), death of a family member (Goenjian et al., 2001), and feeling that one's own life or the lives of family members were in danger (Thienkrua et al., 2006).

Impact of Hurricane Katrina on Children

Research data are becoming available on the effects of Hurricane Katrina on children and adolescents, and studies reflect the unprecedented scale of the storm and complexity of the recovery. In a sample of 166 students in the 9th through 12th grades, Marsee (2008) found that 63% of students had symptoms of PTSD 15–18 months following the hurricane and that the PTSD symptoms together with high levels of aggression were associated with emotional dysregulation. Weems et al. (2009) expanded the understanding of post-Katrina symptomatology finding that, among 52 children with a mean age of 11 years in the 6–7 months following Katrina, the level of posttraumatic stress symptomatology was related to hurricane exposure, female gender, and level of predisaster anxiety. Further research by H. J. Osofsky et al. (2009) found that variables including separation from a caregiver and evacuating to a shelter were associated with posttraumatic stress symptoms in 7- to 19-year-old students 2 years after Hurricane Katrina.

Scheeringa and Zeanah (2008) studied PTSD in 3- to 6-year-old children and found that hurricane-related PTSD was associated with caregiver's level of symptomatology. Spell et al. (2008), in a study of 8- to16-year-old students, found similar results, reflecting the importance of caregiver's symptomatology in predicting a child's psychological status. Terranova, Boxer, and Morris (2009) provided further support for the importance of relationships in their study of sixth grade students who evacuated for Hurricane Katrina but did not experience significant flooding. In this group of children, negative peer interactions were associated with symptoms of PTSD 8 months following the storm. Based on the newly emerging literature, PTSD following Hurricane Katrina appears common, and children's responses are associated with hurricane exposure as well as environmental and relational factors.

Impact of Age on Children's Psychological Responses to Natural Disasters

Some researchers have reported a significant association between the age of a child and the occurrence of psychopathology following disasters. For example, McDermott and Palmer (2002) reported that depression following trauma is more common in younger children than older children in a sample of 8- to 19-year-olds who experienced loss associated with a bushfire. Similarly, Shannon, Lonigan, Finch, and Taylor (1994) reported that in a sample of 5th through 11th graders impacted by Hurricane Hugo, symptoms of posttraumatic stress are more common in younger children than older children. Although these studies suggest that younger children are more susceptible to significant psychopathology following a natural disaster, some studies have indicated otherwise, and a consistent pattern of age differences in response to disasters has not yet been established (Vogel & Vernberg, 1993). A 2008 study of 8- to 15-year-old children found high levels of PTSD 3 years following an earthquake; however, child age did not impact the presence or severity of the disaster response (Bal, 2008). Another study of adolescents in the 9th through 12th grades following Hurricane Katrina demonstrated similar results indicating that age does not necessarily influence traumatic response (Marsee, 2008).

In contrast, research regarding how children express traumatic responses has been well established. Commonly observed traumatic reactions in school-age children include: specific fears, separation difficulties, sleep problems, reenactment of the trauma in play, regression, somatic complaints, irritability, decline in academic performance, fear of recurrence of the trauma, and trauma-related guilt (Steinberg, Brymer, Decker, & Pynoos, 2004; Vogel & Vernberg, 1993). Adolescents, on the contrary, express difficulties through individuation and identity development processes. Pynoos (1993, p. 222) stated that "a trauma-induced sense of discontinuity can give a disrupting influence on the adolescent task of integrating past, present, and future expectations into a lasting sense of identity." Pynoos reaffirmed Blos (1967) who described that adolescence is a period of individuation and elaborated that any threat to this process, through trauma, can potentially disrupt the developmental focus of this important period.

Impact of Gender on Children's Psychological Responses to Natural Disasters

Research on children and adolescents has generally also shown a relation between gender and reports of posttraumatic stress symptoms following exposure to natural disasters. The majority of studies have reported that females are more likely than males to develop PTSD symptoms (Bal, 2008; Giannopoulou et al., 2006; John et al., 2007; Pfefferbaum, 1997; Shannon et al., 1994). For example, following Hurricane Floyd, 9- to 12-year-old females were twice as likely as males to report symptoms of PTSD (Russoniello et al., 2002). Although much of the literature has described gender differences in psychological response to trauma, not all research is consistent on this finding. For example, Kar and Bastia (2006) reported no significant differences in depression and PTSD diagnoses for males and females in high school students following a supercyclone; however, they did report differences in the expression of symptomatology with girls being more likely to report guilt and boys more likely to have increased worry, anhedonia, concentration problems, and academic problems. Shannon et al. (1994) also described gender differences in expression of symptomatology in 9- to 19-year-old children and adolescents with females being more likely to report symptoms associated with emotional reactions and males more often showing symptoms associated with cognitive and behavioral reactions.

Effects of Natural Disasters Over Time

Research on the longitudinal effects of PTSD and depressive symptomatology following disasters indicates that, for the majority of children with PTSD or depressive symptoms, symptoms naturally decrease or remit in the months and years following the event. Vogel and Vernberg (1993) reported that following disasters, symptoms decreased rapidly, usually by 9–14 months postdisaster, and remitted within 18 months to 3 years. However, these authors noted that this recovery may not occur when there has been a severe threat to life or when there is long-term family or community disruption. In children for whom this is the case, persistence or even increases in PTSD and depression have been well documented. McFarlane, Policansky, and Irwin (1987) reported that for 5- to 12-year-old children who witnessed a bushfire, depressive symptoms increased between 2 and 8 months following the disaster and remained elevated when assessed 2 years later. Bal (2008) found that 3 years after an earthquake, high levels of PTSD among 8- to 15-year-old children were associated with chronic stress due to the destruction of housing and community infrastructure. Other studies have found that postdisaster symptoms may have a late onset. For example, John et al. (2007) found that 11% of 5- to 18-year-old children who survived an earthquake and tsunami in Asia developed delayed onset symptoms of PTSD.

Related to hurricane-specific trauma, La Greca et al. (1996) assessed PTSD symptoms in third- through fifth-grade children 3, 7, and 10 months following Hurricane Andrew and found that PTSD symptoms declined over time, although many children remained symptomatic 10 months postdisaster. Symptomatology 10 months postdisaster was related to level of exposure, demographic characteristics, other life stressors, availability of social supports, and type of coping strategies. Halpern and Tramontin (2007) commented that 40% of children who exhibited PTSD symptomatology 10 months after Hurricane Andrew remained symptomatic 3½ years later.

Bokszczanin (2007) highlighted the long-term effects of natural disasters as well as the influence of age and gender on psychological sequelae. Twenty-eight months following a flood, 18% of children ages 11 through 21 met criteria for PTSD. In this sample, the levels of PTSD symptoms were higher in both younger and older females and in younger, but not older males. The author concluded that younger children, regardless of gender, and females, regardless of age, are at higher risk for developing PTSD symptoms following a disaster; he further stated that older males are at the lowest risk of developing PTSD following a disaster.

Few studies have documented continuing symptoms 5 years or more following natural disasters. On such study by Goenjian et al. (2005) followed children after an earthquake and found that PTSD symptoms remained elevated when measured in untreated 16- to 17-year-old children at 18 months and 5 years following the disaster. In another study, data collected 2 years and again at 17 years following the Buffalo Creek Dam collapse indicated that among children who were between the ages of 2 and 15 at the time of the disaster, anxiety, belligerence, somatic concerns, and agitation decreased over time and that, overall, the majority of the sample appeared to have recovered when they were assessed 17 years later. The affected sample no longer showed an elevated level of PTSD when compared to the comparison sample with the level of PTSD decreasing from 32% to 7%. However, females were more likely than males to exhibit long-term symptoms of PTSD, and substance abuse and suicidal ideation were higher among those individuals affected by the dam collapse when compared with individuals in the comparison group (Green et al., 1994).

Overall, the literature indicates that trauma symptoms in children decrease over time (Green et al., 1994; La Greca et al., 1996; Vogel & Vernberg, 1993); however, several researchers have noted that, for some children, trauma symptoms persist (e.g., Bal, 2008; Goenjian et al., 2005; McFarlane et al., 1987). Pynoos (1993) summarized common factors associated with poor long-term outcomes following trauma exposure. He noted that disasters followed by extended periods of high cumulative adversity related to breakdown of infrastructure, ongoing economic consequences, family stress, loss of life and property, and other aspects of slow recovery pose a much greater risk to recovery and well-being.

Patterns of Recovery

Although it is important to document and follow the development and course of symptomatology following natural disasters, this approach does not fully account for the complexities of recovery patterns. Compas, Hinden, and Gerhardt (1995) highlight the importance of addressing the complex pathways to outcomes rather than individual risk or protective factors. Complex pathways include both immediate and chronic consequences of trauma as well as characteristics of the individual and the individual's social environment. Sroufe (1997) expands on this notion, describing the complexities of development following adversity by stating that pathways, including resilience and maladaptation, are not static; rather, they evolve over time and are based on previous experiences and coping capacities. Addressing complex trajectories provides an understanding of positive responses to disasters in addition to the pathological symptomatology discussed previously.

Recently, researchers including Masten and Obradovic (2008) and Bonanno and Mancini (Bonanno, 2004; Bonanno & Mancini, 2008) have focused theoretical discussions on the description of patterns of recovery following trauma. Masten and Obradovic (2008) discuss eight possible outcome trajectories in the context of disaster: *Resistance* refers to the pattern of functioning well during and following a disaster whereas *normal response and recovery* refers to a pattern of initial adaptive problems followed by recovery to good adaptive function. Maladaptive response patterns include *delayed breakdown*, which refers to resistance followed by breakdown in functioning, and *breakdown without recovery*, which refers to a maladaptive response with no improvement. Other patterns of recovery are based on the individual's predisaster functioning. Positive transformation, from either higher or lower levels of predisaster adaptive functioning, may occur following a trauma, and maladaptive patterns of predisaster functioning may continue or worsen following a disaster.

Bonanno and Mancini (2008) have discussed similar recovery patterns, which are described as resilience and recovery patterns as well as chronic and delayed dysfunction patterns. In this conceptualization, *resilience* is described as the absence of a negative response following a trauma, *recovery* is described as having an initial negative response following a trauma but recovering to adaptive functioning, *chronic dysfunction* is described as a negative trauma response that does not abate with time, and *delayed dysfunction* is described as trauma symptomatology that increases over time (Bonanno, 2004; Bonanno & Mancini, 2008).

Regardless of whether a child demonstrates resilience or maladaptation, patterns of recovery must be considered within a developmental framework and address the multiple developmental tasks that children face (Bonanno & Mancini, 2008). Pynoos et al. (1999) advocate a ''complex developmental perspective'' (p. 1542) when considering child traumatic stress and outcomes. The complexity refers to the multiple influences of child factors such as age, gender, previous experiences, and environment (Pynoos et al., 1999). The study of both positive and negative outcomes following disasters can provide more understanding of both traumatic responses and the ways that disruptions of multiple systems impact a child's development following adversity.

Purpose of Study

This study assesses the responses and symptomatology of children and adolescents from a heavily devastated area 2–3 years following Hurricane Katrina and categorizes their responses according to Masten and Obradovic's (2008) theoretical patterns of recovery. More specifically, this study focuses on the developmental factors of age, gender, and life stressors associated with multiple developmental systems and their impact on the course of post-Katrina recovery patterns. The first hypothesis is that student reports of PTSD and depressive symptoms will decrease over time and that changes in symptoms will be related to gender and age. Specifically, it is hypothesized that older children and males will be more likely to report reduction in symptomatology over time than younger children and females. The second hypothesis is that adversity and perceived life stressors will affect the classification of students' recovery patterns into one of four outcome trajectories—stress resistant, normal response and recovery, delayed breakdown, and breakdown without recovery. The third hypothesis is that age and gender will moderate the relation between the life stressor variables and outcome trajectories.

Method

As part of a larger assessment of children's overall needs after Hurricane Katrina, the current study focuses on data collected in St. Bernard Parish for the purpose of guiding the delivery of intervention services. The St. Bernard Parish School District administered the National Child Traumatic Stress Network (NCTSN) Hurricane and Assessment Referral Tool (NCTSN, 2005) to students in the schools, and faculty from Louisiana State University Health Sciences Center (LSUHSC) Department of Psychiatry was available to answer questions or provide support. Screeners were administered to students in the target community if they were in 4th through 12th grades; there were no major exclusion criteria within this group of students. The assessment was administered confidentially, but not anonymously, so that it was possible to provide further evaluation for students who scored above the cutoff for mental health symptoms and respond to those who requested a meeting with a counselor. Students were given the opportunity to refuse participation on all or parts of the screen at their discretion. Reports, including compilations of students' responses, were given to the schools in order to provide them with more information about student's concerns in an effort to improve services and enhance the school environment. This study was submitted and received approval by the IRB Board at LSUHSC.

Students were assessed each school year for mental health needs beginning immediately after Hurricane Katrina (2005–2006) and continuing to the most recent 2007–2008 school year. Given the variability of the students returning during the 1st year and the influx of students in subsequent years, Years 2 and 3 were used in the analysis. Among the total group of students, we had an approximately 67% rate of response: 1,826 surveys were returned for the 2006–2007 academic year and 1,380 were returned in the 2007–2008 academic year. Students were selected for inclusion in the current study if they were screened during both the 2nd and 3rd post-Hurricane Katrina years (n = 542). Acock (2005) indicates that attrition is a major concern in all longitudinal community-based surveys, and a 50% rate of attrition over time is common. Although the rate of attrition in the current study is higher, it is not unexpected in a post-disaster environment. Mobility rates and school absences throughout this time were high, with difficulties in housing, family illnesses or stressors, parent employment, and families moving to other locations as community recovery languished. Missing data analyses indicate that these 542 students do not differ significantly on whether or not students met the cutoff score for enhanced mental health services at Time 3 when compared with the group of 1,258, $\chi^2(1) = 1.67$, *ns*. Among those 542 students, 387 students (71.4%) provided complete data for the variables related to the research questions in the current study. Furthermore, missing data analyses indicate that the 387 students do not differ significantly from the students without complete data on all outcome variables: outcome trajectories, $\chi^2(1) = 4.22$, *ns*; depression, $t(504) = 0.09$, *ns*; PTSD, $t(512) = 0.12$, *ns*; re-experiencing, $t(532) = 0.47$, *ns*; avoidance, $t(533) = 0.24$, *ns*; and hyper-arousal, $t(527) = 0.19$, *ns*.

Participants

Three hundred eighty-seven students with complete data for the variables were included in the data analysis: 176 (45.5%) were male, and 211 (54.5%) were female. The mean age of the students was 14.27 years (*SD* = 2.13). Age was further categorized into three age groupings that correspond to the different developmental tasks and competencies of middle childhood (9–11), early adolescence (12–14), and middle adolescence (15–18). Forty-four (11.4%) of the students were in the 9–11 age group (male = 17, female = 27), 165 (42.6%) were in the 12–14 age group (male = 77, female = 88), and 178 (46.0%) were in the 15–18 age group (male = 82, female = 96). Two hundred eighty-two (72.9%) of the students identified themselves as Caucasian, 47 (12.1%) identified as African American, 20 (5.2%) identified as Hispanic, 19 (4.9%) identified as mixed race, and 19 (4.9%) reported other or unknown ethnic affiliation.

Measures

The measure used for this study was the NCTSN Hurricane Assessment and Referral Tool for Children and Adolescents (NCTSN, 2005). The NCTSN assessment scale of posttraumatic symptoms was based on the validated UCLA PTSD Index and includes a number of questions regarding depressive symptoms. The measure consists of 22 items on a 4-point Likert scale, with 1 representing *not at all*, 2 representing *a little bit*, 3 representing *quite a bit*, and 4 representing *very much*. The measure was modified by the LSUHSC Department of Psychiatry in collaboration with school administrators and students to aid in the students' ability to complete the scale. In addition to modifications of the wording of the screener to make it easier for the students to fill out themselves, another modification included changing the original 5-point Likert scale to a 4-point Likert scale. To score the screener, 1 point is given for each question rated 3 or 4; four or more scores of 3 or higher suggest the need for referral for further evaluation and enhanced mental health services.

Symptomatology. Depression, overall PTSD, and PTSD subscales symptoms (re-experiencing, avoidance, and hyper-arousal) were derived from the NCTSN screener using clinical classification based on *DSM–IV–TR* criteria (American Psychiatric Association, 2000). The internal consistencies (α coefficient) for each scale and subscale were as follows: .91 for depression and .90 for overall PTSD. PTSD subscale alphas were .75 for re-experiencing, .60 for avoidance, and .81 for hyper-arousal. Depressive symptoms included: 6–9, 11–18, and 20; for example, nothing is fun anymore (Item 6), sleep difficulties (Item 7), feeling jumpy nervous (Item 8), difficulty concentrating (Item 9), feeling irritable (Item 11), feeling sad (Item 12), appetite change (Item 13), headaches or stomachaches (Item 14), less energy (Item 15), difficulty with schoolwork (Item 16), worry about something bad happening (Item 17), difficulty with family or friends (Item 18), and harder to enjoy activities (Item 20). PTSD symptoms included: upset when thinking about hurricane (1), nightmares (2), upsetting thoughts (3), try not to think or talk about hurricane (4), stay away of trauma reminders (5), sleep difficulties (7), feeling jumpy or nervous (8), difficulty concentrating (9), worry about what will happen (10), difficulty with schoolwork (16), and worry about something bad happening (17). PTSD subscale symptoms included items: 2, 3, and 10 for re-experiencing; 1, 4, and 5 for avoidance; and 7–9 and 16 for hyper-arousal. These scores represent symptoms based on a self-report measure and are not considered diagnoses of PTSD or depression.

Outcome trajectories. For person-centered analyses, students were classified according to recovery patterns adapted from Masten and Obradovic (2008). The trajectories included stress resistant, normal response and recovery, delayed breakdown, and breakdown without recovery. Classification was based on whether or not participants met the criteria for referral in the 2nd (2006–2007) and 3rd (2007–2008) post-Katrina school years. Students who did not meet the cutoff for referral on the screening measure in either year were classified as stress resistant. Students who met the cutoff on the screening measure in the second post-Hurricane Katrina year but did not meet the cutoff in the 3rd year were classified as normal response and recovery. Students who met the cutoff for both post-Hurricane Katrina years were classified as breakdown without recovery, and students who did not meet the cutoff in the 2nd post-Hurricane Katrina year, but met the cutoff in the 3rd year were classified as delayed breakdown.

Life stressors. Adversity scores included recovery stressors related to the aftermath of the hurricane, significant loss or trauma pre- and post-Hurricane Katrina, and current perceptions of life problems. Hurricane recovery stressors were assessed through student endorsement of the following variables: consultation with a mental health counselor or therapist since the storm, housing others due to storm, still living in a trailer, home needed renovation, currently residing in home under renovation, and parent still unemployed due to hurricane. Hurricane recovery stressors were also assessed by students' reports on a 1–10 scale of how connected they felt to their family, with 1 representing limited or no connectedness; total number of places they reported living since the storm, and the total number of schools they attended since the storm. Significant loss or trauma was operationalized as having experienced a major loss or trauma prior to or since Hurricane Katrina. The index of current problems was based on concerns reported in the 3rd-year assessment. Problems were operationalized as dichotomous variables on whether or not the student reported family, school, friend, financial, housing, or whether they reported problems with neighborhood violence. Descriptive statistics on life stressor variables by outcome trajectory are presented later in Table 3.

Results

Reported Symptoms of Depression and PTSD

Mixed method (two-between/one-within) analysis of variance (ANOVA) was used to assess

Table 1
Mixed Method ANOVAs on Symptoms by Age and by Gender

Source	*df*	*F*	η^2
Between subjects			
Depression			
Age	2	8.45*	.043
Gender	1	8.67*	.022
Age × Gender	2	0.27	.001
	381	(0.64)	
PTSD			
Age	2	9.40*	.047
Gender	1	14.14*	.036
Age × Gender	2	0.14	.001
	381	(0.63)	
Re-experiencing			
Age	2	8.92*	.045
Gender	1	16.12*	.041
Age × Gender	2	0.15	.001
	381	(0.74)	
Avoidance			
Age	2	5.60*	.029
Gender	1	21.07*	.052
Age × Gender	2	1.74	.009
	381	(0.68)	
Hyper-arousal			
Age	2	9.19*	.825
Gender	1	3.32	.046
Age × Gender	2	0.40	.002
	381	(0.77)	
Within subjects			
Depression	1	81.03*	.175
		(0.16)	
PTSD	1	116.36*	.234
		(0.17)	
Re-experiencing	1	111.55*	.226
		(0.26)	
Avoidance	1	48.94*	.114
		(0.34)	
Hyper-arousal	1	50.44*	.117
		(0.24)	

Note. Numbers in parentheses represent mean squared error. Depression = D, PTSD = P, re-experiencing = R, avoidance = A, and hyper-arousal = H. 2006–2007 > 2007–2008; for 2006–2007 and 2007–2008, 9–11 > 12–14 on D, P, R, and H, and 9–11 > 15–18 on D, P, R, and H; for 2007–2008, 9–11 > 15–18 on A. Females > males on D, P, R, and A. ANOVA = analysis of variance; PTSD = posttraumatic stress disorder.
$^*p < .05$.

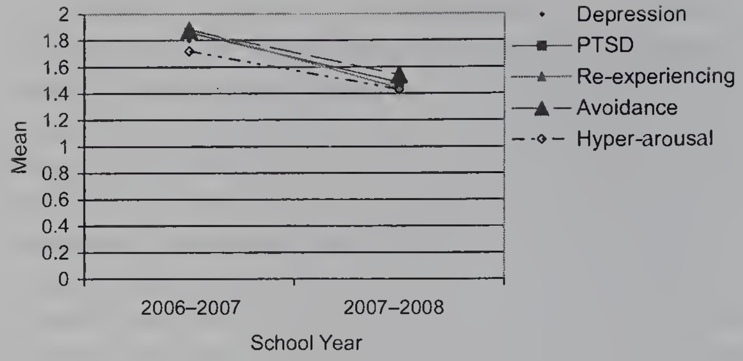


Figure 1. Line graph displaying symptom means by school year.

differences between the 2nd and 3rd post-Katrina years (2006–2007 vs. 2007–2008) on students' reported symptoms of depression and PTSD by age (9–11 vs. 12–14 vs. 15–18) and by gender (male vs. female). Table 1 presents the results, where within-subjects main effects were shown on depression, overall PTSD, and PTSD subscales (re-experiencing, avoidance, and hyper-arousal). Within-subjects main effects showed a statistically significant decrease on symptom scores (see Figure 1) between the 2nd and 3rd post-Katrina years (2006–2007 vs. 2007–2008). Between-subjects main effects were shown on depression, PTSD, re-experiencing, avoidance, and, hyper-arousal by age, and on depression, PTSD, re-experiencing, and avoidance by gender. No significant Age × Gender interaction effects were revealed. Table 2 presents the mean symptom scores by age and by gender. Interpretations of results using post hoc comparisons for the significant between subject main effects are presented next; for all post hoc comparisons alpha levels were set at $p < .01$.

Age. Fifteen post hoc tests were conducted to assess if differences exist from the 2nd to 3rd post-Katrina years on the symptoms (depression, PTSD, re-experiencing, avoidance, and hyper-arousal) for each age category. Results of the paired sample *t* tests revealed statistically significant decreases on symptoms from the 2nd to 3rd post-Katrina years for students in the: 9–11, $t(43) = 2.87–4.66$, *p*s < .01; 12–14, $t(164) = 4.80–9.03$, *p*s < .01; and 15–18, $t(177) = 4.72–7.81$, *p*s < .01, age categories. Ten one-way ANOVAs were also conducted to assess age differences on the 2006–2007 and the 2007–2008 symptoms measures (depression, PTSD, re-experiencing, avoidance, and hyper-arousal). Significant age differences were revealed on the 2006–2007 symptom measures of depression—PTSD, re-experiencing, and hyper-arousal, $F(2, 384) = 4.86–8.52$, *p*s < .01—and on the 2007–2008 symptom measures—depression, PTSD, re-experiencing, avoidance, and hyper-arousal, $F(2, 384) = 4.86–8.32$, *p*s < .01. Scheffe post hoc tests suggest that for both academic years (2006–2007 and 2007–2008), students in the 9–11 age category scored significantly higher on depression, PTSD, re-experiencing, and hyper-arousal compared with students in the 12–14 and 15–18 age categories. Tests also suggest

Table 2
Means Symptom Scores by Age and by Gender

		Ages 9–11			Ages 12–14			Ages 15–18			Total		
Symptom	Time	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
Depression	1	2.07	2.20	2.15	1.65	1.93	1.8	1.57	1.86	1.72	1.65	1.93	1.80
	2	1.74	1.87	1.82	1.33	1.52	1.43	1.31	1.56	1.45	1.36	1.58	1.48
PTSD	1	2.15	2.32	2.25	1.7	1.95	1.83	1.6	1.94	1.78	1.70	1.99	1.86
	2	1.61	1.95	1.82	1.33	1.57	1.46	1.28	1.55	1.42	1.33	1.61	1.48
Re-experiencing	1	2.16	2.46	2.34	1.71	2.01	1.87	1.64	1.93	1.80	1.72	2.03	1.89
	2	1.51	1.96	1.79	1.29	1.52	1.41	1.24	1.52	1.39	1.29	1.57	1.44
Avoidance	1	2.00	2.33	2.2	1.77	1.93	1.86	1.54	2.07	1.83	1.69	2.05	1.88
	2	1.53	1.96	1.8	1.43	1.67	1.56	1.28	1.60	1.45	1.37	1.68	1.54
Hyper-arousal	1	2.15	2.12	2.13	1.58	1.79	1.69	1.52	1.74	1.64	1.61	1.81	1.72
	2	1.74	1.81	1.78	1.28	1.48	1.38	1.28	1.47	1.38	1.32	1.52	1.43

Note. Time 1 = 2006–2007 school year; Time 2 = 2007–2008 school year. Depression = D, PTSD = P, re-experiencing = R, avoidance = A, and hyper-arousal = H. 2006–2007 > 2007–2008; 9–11 > 12–14 on D, P, R, A, and H, and 9–11 > 15–18 on D, P, R, and H. Females > males on D, P, R, and A. PTSD = posttraumatic stress disorder.

that for the 2007–2008 academic year, students in the 9–11 age category scored significantly higher on avoidance compared with students in the 15–18 age category. There were no significant differences on symptom measures between the 12–14 and the 15–18 age categories.

Gender. Eight post hoc tests were conducted to assess whether differences exist between the 2nd and 3rd post-Katrina years on the symptoms (depression, PTSD, re-experiencing, and avoidance) for each within-gender subgroup. Results of the paired sample t tests revealed statistically significant decreases on symptoms between the 2nd and 3rd post-Katrina years for both males, $t(175) = 5.58–8.57$, $ps < .001$, and females, $t(210) = 6.14–9.37$, $ps < .001$. Eight independent sample t tests were also conducted to assess gender differences on the 2006–2007 and the 2007–2008 symptom measures of depression, PTSD, re-experiencing, and avoidance. These post hoc tests further revealed that females scored significantly higher than males on the 2006–2007 measures of depression, PTSD, re-experiencing, and avoidance, $t(385) = 3.87–4.53$, $ps < .001$, and the 2007–2008 measures of depression, PTSD, re-experiencing, and avoidance, $t(385) = 3.83–4.57$, $ps < .001$.

Outcome Trajectories Following Trauma

Students who completed the screening measure during the 2nd and 3rd post-Katrina years were classified according to the outcome trajectories described by Masten and Obradovic (2008): 175 (45.2%) students were classified as stress resistant, 105 (27.1%) as normal response and recovery, 18 (4.7%) as delayed breakdown, and 89 (23.0%) as breakdown without recovery. Mean total screener scores for the 2007–2008 school year by outcome trajectory are as follows: 0.74 ($SD = 0.83$) stress resistant, 1.49 ($SD = 1.09$) normal response and recovery, 5.39 ($SD = 1.91$) delayed breakdown, and 8.27 ($SD = 4.09$) breakdown without recovery, demonstrating that posttraumatic stress symptoms increased by trajectory classification, from resistant and recovery to delayed and continued breakdown classifications. Descriptive statistics on life stressors, age, and gender by outcome trajectory are presented in Table 3.

The following analyses were conducted to assess the relation between outcome trajectories and life stressors. Given the limited number of students in the delayed breakdown classification, chi-square tests were conducted to assess whether differences existed between the delayed breakdown and breakdown without recovery trajectories by the life stressor variables. Results revealed no significant differences, which supported combining of the two categories. Therefore, for statistical purposes, the delayed breakdown category was collapsed into the breakdown without recovery category ($n = 108$).

Although St. Bernard Parish was a heavily devastated area, analyses were conducted to assess the level of exposure by outcome trajectory. Level of exposure was operationalized by six questions including, did the student: have to be separated from a caregiver (23.0%), witness injury (13.7%), experience personal injury (4.1%), have family or friends injured (18.6%), have family or friends

Table 3

Chi-Square Statistics on Trajectories by Age, Gender, and Life Stressors

Variables	Stress resistant (n = 174)	Normal response and recovery (n = 105)	Breakdown without recovery (n = 108)	χ^2
Age				
9–11	10 (22.7)	12 (27.3)	22 (50.0)	15.79**
12–14	76 (46.1)	50 (30.3)	39 (23.6)	
15–18	88 (49.4)	43 (24.2)	47 (26.4)	
Female	74 (42.5)	65 (61.9)	72 (66.7)	18.83**
Consulted with mental health counselor/therapist since storm	24 (13.8)	17 (16.2)	38 (35.2)	20.35**
Problems/worries				
Family	6 (3.4)	11 (10.5)	27 (25.0)	30.83**
School	23 (13.2)	23 (21.9)	52 (48.1)	43.88**
Friend	12 (6.9)	12 (11.4)	42 (38.9)	51.44**
Financial	10 (5.7)	10 (9.5)	22 (20.4)	14.99**
Housing	5 (2.9)	6 (5.7)	21 (19.4)	25.36**
Neighborhood violence	5 (2.9)	6 (5.7)	17 (15.7)	16.94**
Previous loss or trauma	35 (20.1)	28 (26.7)	42 (38.9)	11.89**
Post hurricane loss or trauma	37 (21.3)	33 (31.4)	47 (43.5)	15.74**
House others due to storm	12 (6.9)	8 (7.6)	19 (17.6)	9.37**
Still living in a trailer	23 (13.2)	18 (17.1)	26 (24.1)	5.49**
House fixed or needed renovation	130 (74.7)	79 (75.2)	75 (69.4)	1.20
Currently renovating house	38 (21.8)	26 (24.8)	32 (29.6)	2.17
Parent still unemployed	18 (10.3)	13 (12.4)	21 (19.4)	4.88

Note. Numbers in parentheses represent percentages.
**$p < .01$.

killed (12.7%), and have personal belongings destroyed (94.8%). Questions were summed together to obtain a total level of exposure score; the minimum score was 0 and the maximum score was 5 (M = 1.67, SD = 0.84). An ANOVA was conducted, $F(2, 384) = 9.49$, $p < .001$, and revealed that differences existed among the outcome trajectories and level of exposure. Scheffe post hoc tests were conducted and revealed that students in the stress-resistant trajectory had lower levels of exposure (M = 1.47, SD = 0.84) compared with students in the normal response and recovery (M = 1.72, SD = 0.89) and breakdown without recovery (M = 1.94, SD = 1.03) trajectories.

Bivariate Analyses

Chi-square tests were conducted to assess relation between outcome trajectories (stress resistant, normal response and recovery, and breakdown without recovery) and age categories, gender, and life stressors. Results and descriptive statistics are presented in Table 3. Post hoc examination was also conducted to assess individual differences. Results revealed significant age differences, suggesting that a greater proportion of students in the breakdown without recovery group were in the 9–11 age category and lower proportions were in the 12–14 age category compared with the stress-resistant and normal response and recovery trajectories. Results on age differences also revealed that a greater proportion of students in the breakdown without recovery group were in the 9–11 age category and lower proportions were in the 15–18 age category compared with the stress resistant trajectory. Results on gender differences were significant indicating that a greater proportion of females were classified in the breakdown without recovery and normal response and recovery group compared with the stress-resistant group.

Results indicate that greater proportions of students in the breakdown without recovery classification reported consulting with a mental health counselor or therapist; endorsing family, school, friend, financial, housing, or neighborhood violence problems or worries; experiencing previous loss or trauma; posthurricane major loss or trauma; housing others due to the storm; and still living in a trailer when compared to students with stress-resistant and normal response and recovery classifications. The results further indicate that greater proportions of students in the normal response and recovery group endorsed school or friend problems and experienced previous loss or trauma or posthurricane loss or trauma, compared with the stress-resistant group. There were no significant differences among outcome trajectories and house-needed renovation, currently residing in house under renovation, and parent still unemployed due to hurricane.

One-way ANOVAs were conducted to assess if differences existed on family connectedness, number of places lived since the storm, and number of schools attended since the storm by outcome trajectory (stress resistant, normal response and recovery, and breakdown without recovery). Results and descriptive statistics are presented in Table 4. Scheffe post hoc tests revealed that students in the breakdown without recovery group reported lower family connectedness, compared with students in the stress-resistant and normal response and recovery groups. Results also indicate that students in the breakdown without recovery group reported living in more places since the storm, compared with students in the stress-resistant group. There were no significant differences in outcome trajectory by number of schools attended since the storm.

Multinomial Logistic Regression

A multinomial logistic regression was conducted to assess the additive effect of age, gender, and the life stressor variables on classification of the outcome trajectories (stress resistant vs. normal response and recovery vs. breakdown without recovery). The following life stressor variables were included in the regression due to significant bivariate relation with the outcome trajectories: consultation with a mental health counselor or therapist since the storm; endorsement of family, school, friend, financial, housing, or neighborhood violence problems; prehurricane losses or trauma; posthurricane losses or trauma; housing others due to storm; still living in a trailer; family connectedness; and number of places lived since the storm. Breakdown without recovery was the reference category and the enter method was used. All variance inflation factors were under 2 suggesting the absence of multicollinearity; all other assumptions were met.

There was a good model fit (discrimination among groups) on the basis of the constant only model (no predictors were entered and assessed if classification was based on chance alone) $\chi^2(640.80)$, $p = .72$. After adding the life stressor variables, the test of the overall model, likelihood ratio, was significant, $\chi^2(32) = 143.32$, $p < .001$, Nagelkerke = .351. The life stressor variables collectively classified 62.8% of the students as stress resistant, 12.4% as normal response and recovery, and 24.8% as the breakdown without recovery with an overall correct classification of 57.9%.

Table 5 shows the contribution of the individual predictors to the overall model by comparing models with and without each predictor. Age significantly enhanced prediction, with students ages 15–18 compared to younger students (ages 9–11) being 4.11 times more likely to be in the stress-resistant trajectory than the breakdown without recovery trajectory. Gender also significantly enhanced prediction, with female students being almost twice as likely (OR = 1.97, $p < .05$) to be classified as breakdown without recovery compared with stress resistant. Three of the life stressor variables significantly enhanced prediction. The

Table 4
ANOVAs on Outcome Trajectories and Life Stressors

	Outcome trajectories							
	Stress resistant (n = 174)		Normal response and recovery (n = 105)		Breakdown without recovery (n = 108)			
Life stressor variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2
Family connectedness	7.87	2.31	8.15	2.18	6.85	2.81	8.79*	.043
Number of places lived since the storm	2.95	1.22	3.17	1.31	3.31	1.15	3.03*	.016
Number of schools attended since storm	2.62	0.93	2.75	1.05	2.89	1.09	2.37	.012

Note. ANOVA = analysis of variance.
*$p < .05$.

Table 5
Logistic Regression Analysis of Outcome Trajectories as a Function of Life Stressors

Variable	Category	*SE*	Exp(*B*)	95% CI	
				Lower	Upper
Age, $\chi^2(2) = 9.65$*	Resistant vs. breakdown				
	9–11 vs. 15–18	0.51	0.24*	0.09	0.67
	12–14 vs. 15–18	0.33	0.98	0.52	1.85
	Normal vs. breakdown				
	9–11 vs. 15–18	0.49	0.44	0.17	1.15
	12–14 vs. 15–18	0.34	1.24	0.64	2.40
Gender, $\chi^2(2) = 8.59$*	Resistant vs. breakdown	0.31	0.51*	0.28	0.93
	Normal vs. breakdown	0.32	1.00	0.53	1.88
Consulted with mental health counselor or therapist since storm, $\chi^2(2) = 6.53$*	Resistant vs. breakdown	0.37	0.44*	0.21	0.90
	Normal vs. breakdown	0.38	0.43*	0.21	0.91
Family problems, $\chi^2(2) = 12.53$*	Resistant vs. breakdown	0.44	0.23*	0.10	0.55
	Normal vs. breakdown	0.43	0.36*	0.16	0.82
School problems, $\chi^2(2) = 11.07$*	Resistant vs. breakdown	0.36	0.31*	0.15	0.62
	Normal vs. breakdown	0.35	0.50*	0.25	0.99
Friend problems, $\chi^2(2) = 3.56$	Resistant vs. breakdown	0.55	0.38	0.13	1.11
	Normal vs. breakdown	0.46	0.80	0.32	1.97
Financial problems, $\chi^2(2) = 0.89$	Resistant vs. breakdown	0.56	0.59	0.20	1.78
	Normal vs. breakdown	0.54	0.82	0.28	2.37
Housing problems, $\chi^2(2) = 3.19$	Resistant vs. breakdown	0.71	0.36	0.09	1.44
	Normal vs. breakdown	0.66	0.38	0.10	1.38
Neighborhood violence problems, $\chi^2(2) = 2.78$	Resistant vs. breakdown	0.63	0.36	0.10	1.24
	Normal vs. breakdown	0.56	0.65	0.22	1.96
Previous loss or trauma, $\chi^2(2) = 1.02$	Resistant vs. breakdown	0.37	0.69	0.34	1.42
	Normal vs. breakdown	0.37	0.78	0.38	1.61
Post hurricane major loss or trauma, $\chi^2(2) = 1.40$	Resistant vs. breakdown	0.35	0.75	0.38	1.49
	Normal vs. breakdown	0.35	1.06	0.53	2.10
House others due to storm, $\chi^2(2) = 3.13$	Resistant vs. breakdown	0.49	0.46	0.17	1.20
	Normal vs. breakdown	0.50	0.47	0.17	1.25
Still living in a trailer, $\chi^2(2) = 0.93$	Resistant vs. breakdown	0.41	0.67	0.30	1.50
	Normal vs. breakdown	0.41	0.80	0.36	1.80
Family connectedness, $\chi^2(2) = 3.22$	Resistant vs. breakdown	0.07	1.06	0.93	1.21
	Normal vs. breakdown	0.07	1.13	0.99	1.30
Number of places lived since the storm, $\chi^2(2) = 1.00$	Resistant vs. breakdown	0.13	0.99	0.77	1.28
	Normal vs. breakdown	0.13	1.10	0.85	1.43

*$p < .05$.

greatest effect in classification of the outcome trajectories was contributed by the following variables: needed to consult with a mental health professional due to stress since the storm and endorsement of family and school problems. The following paragraph presents the interpretation of the odds ratios for the significant individual predictors.

Students who needed to consult with a mental health professional had 2.30 greater odds ($p < .05$) of being classified as breakdown without recovery compared with stress resistant and 2.31 greater odds ($p < .05$) of being classified as breakdown without recovery compared with normal response and recovery. Students who endorsed family (OR = 4.33, $p < .001$) or school (OR = 3.25, $p < .001$) problems were over 3 times more likely to be in the breakdown without recovery trajectory than the stress-resistant trajectory. Students who endorsed family (OR = 2.80, $p < .05$) or school (OR = 2.01, $p < .05$) problems were twice as likely to be in the breakdown without recovery trajectory compared to normal response and recovery.

Moderation analyses were also conducted to assess whether interactions existed among age, gender, and the three variables with significant odds ratios in the regression model (needing to consult with a mental health professional since the storm, and endorsing family and school problems). All variables were centered prior to analysis and mean age was used. Results showed that the interactions of age with mental health consultation and endorsing family or school problems were not significant. Results also showed that the interactions of gender with mental health consultation and endorsing family or school problems were not significant. The lack of moderation suggests that age and gender do not alter the relation between mental health consultation and endorsing family or school problems with the outcome trajectory classifications.

Summary

The first hypothesis, that student reports of PTSD and depressive symptoms would decrease over time and changes in symptoms would be related to age and gender, was supported. Results showed a statistically significant decrease in depression, overall PTSD, and PTSD subscales (re-experiencing, avoidance, and hyper-arousal) scores between the 2nd and 3rd post-Katrina years (2006–2007 vs. 2007–2008). Results also revealed that younger students (ages 9–11) scored significantly higher on depression, PTSD, re-experiencing, avoidance, and hyper-arousal compared with older students (ages 12–18). Results also showed that females scored significantly higher on depression, PTSD, re-experiencing, and avoidance than males.

The second hypothesis, that adversity and perceived life stressors would classify students' recovery patterns into one of four outcome trajectories (stress resistant, normal response and recovery, breakdown without recovery), was also supported. The greatest influence in classification of the outcome trajectories was contributed by the following variables: age, gender, consultation with a mental health professional since the storm, and endorsement of family or school problems. Although bivariate relation existed between outcome trajectories and endorsement of friend, financial, housing, or neighborhood violence problems; prehurricane losses or trauma; posthurricane losses or trauma; housing others due to storm; still living in a trailer; and number of places lived since the storm, they did not contribute to the overall regression model. The life stressor variables, home needed renovation, currently residing in home under renovation, parent still unemployed due to hurricane, and number of schools attended since the storm, were not related to the outcome trajectories.

The third hypothesis, that age and gender would moderate the relation between the life stressor variables and outcome trajectories, was not supported. Results suggest that age and gender do not alter the relation between mental health consultation and endorsing family or school problems with the outcome trajectory classifications.

Discussion

The purpose of this study was to examine not only the symptomatology endorsed by children who lived in an area heavily devastated by the major disaster, Hurricane Katrina, but also to learn more about the recovery patterns of these children and life stressors that influence these patterns. This study focused on the developmental factors of age, gender, and life stressors associated with multiple developmental systems and their impact on the course of children's recovery patterns. In order to track recovery, stressors and resultant symptoms in these youth were studied for 3 years following the hurricane.

Major disasters with slow recovery can result in acute and chronic psychological effects, which can impact development by interrupting normal developmental trajectories (Pynoos et al., 1999). Compas et al. (1995) describe the multiple influences on development including internal processes and external environmental factors, which lead to various adaptive and maladaptive trajectories of development. They assert that, in general, adolescents who live in environments that are relatively free of significant stressors will demonstrate ongoing and stable adaptive functioning characterized by a positive sense of self and a lack of severely problematic behaviors or significant emotional problems. The overall adaptive pathway may veer toward dysfunction when trauma is a precursor to clinical depression and PTSD, which cause significant distress in social, occupational, or other important areas of functioning (American Psychiatric Association, 2000). As adolescent functioning is associated with adult functioning (Compas et al., 1995), trauma in childhood and adolescence may have a long-lasting impact. Due to the long-term life stressors associated with Hurricane Katrina as well as individual and environmental supports, children and adolescents have exhibited complex reactions and have demonstrated trauma responses ranging from stress resistance and recovery to continued breakdown. The results of this study provide insight into factors that contribute to different outcomes.

One of the most important outcomes of this study relates to the strengths shown by students in St. Bernard Parish. Not only did overall symptoms decrease from the 2nd to 3rd post-Katrina years, but also 45.2% of the children and adolescents were classified as stress resistant and did not report high levels of symptomatology when assessed 2 and 3 years following the disaster. These data are consistent with the findings of Bonanno and Mancini (2008) who estimated that about half of all children and adolescents will display no immediate or long-term psychological sequelae following traumas including bereavement, terrorism, and health-related traumas. Further, 27.1% of the children and adolescents were classified as normal response and recovery, which means that they reported symptoms of PTSD and depression during the 1st year following the hurricane, but few symptoms during the 2nd year. The finding that 72.3% of the children were classified as stress resistant or experienced normal response and recovery, although consistent with other literature on disasters (La Greca et al., 1996), is striking given the extent of the devastation, slow recovery, and conditions under which participants lived in St. Bernard Parish following Hurricane Katrina. Perhaps these are the children described as ''resilient'' in the current literature.

Resilience has been a topic of great interest to researchers as a way to potentially explain why some children succeed despite adversity. Although researchers have posited numerous definitions of resilient patterns of recovery including ''succeeding in spite of serious challenges to their development'' (O'Daugherty & Masten, 2005, p. 17), ''early effective efforts at adjustment and recovery'' (Pynoos et al., 1999, p. 1542), ''facing adversity and nonetheless do(ing) well'' (Sroufe, 1997, p. 256), and ''the ability to maintain healthy levels of psychological and physical functioning as well as the capacity for generative experiences and positive emotions'' (Bonanno & Mancini, 2008, p. 371), they generally agree that adaptive patterns of functioning in children following a trauma are common (Bonanno & Mancini, 2008; Masten, 2001). Sroufe, Egeland, Carlson, and Collins (2005) expand the understanding of resilience by describing the concept ''not as an individual trait but a feature of the developmental system'' (p. 227) that is associated with a foundation of early support and attachment. Positive recovery patterns are associated with self-efficacy, positive coping and problem-solving skills, self-regulation, and supportive social systems including family, peer group, and school (Masten & Obradovic, 2008).

Although our findings, in general, suggest a high level of resilience among children, it is also important to note that during the 3rd year after Hurricane Katrina, 27.7% experienced breakdown without recovery or delayed breakdown. In other words, although the majority of the sample showed positive signs of recovery with symptoms decreasing over time, more than one fourth of children reported significant symptoms of posttraumatic stress and depression and met the cutoff to qualify for mental health referral 3 years after the hurricane. These students also reportedly had higher levels of exposure, which may affect their resilience. The elevated levels of long-term symptomatology were not unexpected given the slowness of recovery and conditions of the community; however, this finding is important in understanding recovery in order to learn more about why some children showed very few symptoms and others remained symptomatic over time.

Based on the results of this study, several factors were associated with recovery patterns. Specifically, younger students, females, individuals who consulted with a mental health professional since the storm, experienced previous loss or trauma, experienced posthurricane major loss or trauma, housed others due to the storm, continued to live in a trailer, had limited family connectedness, lived in a greater number of places since the storm, and endorsed financial, housing, family, school, friend, or neighborhood violence problems were more likely to have continued negative responses to the hurricane. The most influential of these developmental and systemic factors related to students' responses were age, gender, consulting with a mental health professional, and endorsing family and school problems. These issues will be elaborated as follows.

In terms of age differences, in general, children who are younger may be more vulnerable as they have not had the opportunity to develop their own coping capacity when compared with older children, and they likely depend much more on adults in their environment for support (Osofsky, 2004). The results of this study were consistent with the idea of increased vulnerability for younger children and revealed that younger students (ages 9–11) scored significantly higher on depression, PTSD, re-experiencing, avoidance, and hyper-arousal compared with older students (aged 12–18). Further, students in the breakdown without recovery category when compared with those in the stress resistant group were significantly younger. These younger children may be less likely to be able to access systems of support to cope with the continuing disruption in their homes and neighborhoods. A study by Cryder, Kilmer, Tedeschi, and Calhoun (2006) supports this outcome as they found that among 6- to 15-year-old children, older children, when compared with younger children, were more likely to have social support systems that included friends and adults outside of their immediate family. Social support was associated with positive outcomes including posttraumatic growth approximately 1 year following Hurricane Floyd. Thus, younger children's limited social support may be considered a risk factor to positive postdisaster recovery. This issue of increased vulnerability for younger children highlights the importance of developing strategies to support the developmental needs of children exposed to trauma, a consideration that is rarely addressed in overall planning for disaster response (Osofsky et al., 2007).

As might be expected and consistent with much of the available literature related to gender effects on mental health (Bal, 2008; Giannopoulou et al., 2006; John et al., 2007; Pfefferbaum, 1997; Shannon et al., 1994), females reported more symptoms of depression than males. A greater proportion of females were classified in the breakdown without recovery and normal response and recovery group when compared with males for whom a greater proportion were in the stress-resistant group. Overall, it appears that the male patterns of functioning over the 2nd and 3rd years following the disaster were more likely to indicate positive patterns of recovery, whereas some females showed recovery to good adaptive functioning and others displayed more maladaptive patterns. These results may be related both to the increased vulnerability of females to depressive symptoms and differences in the availability of support networks that are often so important for recovery. An alternative interpretation for this finding that boys are more likely to be stress resistant than girls is that boys are more likely to underreport symptoms of posttraumatic stress and depression. This is supported by research that has shown that girls and boys respond with different symptomatology following trauma. It is possible that boys respond to trauma with increased behavioral, attention, and academic problems, whereas girls may be more likely to respond to trauma with increased guilt and emotional reactions (Kar & Bastia, 2006; Shannon et al., 1994).

As would be expected, children and adolescents who experienced breakdown without recovery were more likely to have consulted a mental health professional related to stress following the hurricane and to report school and family problems. This finding may be interpreted as a positive indicator that school and community mental health services are available and accessible to children who exhibit long-term symptomatology and need mental health intervention. This may also contribute to the strength of children in this classification in that their families and schools recognize their need for and seek out mental health treatment. The issue of utilization of mental health services highlights the importance of having such supports and services easily accessible, especially in school settings in the postdisaster environment.

Students with concerns about their families were more likely to continue to show symptoms 3 years after the hurricane. This finding highlights the importance of supporting parents and the caregiver–child relationship after disasters and is consistent with the resilience literature (Masten, 2001). The current study focuses on stressors faced by children following Hurricane Katrina; however, their parents also experienced significant distress following the storm (Kessler et al., 2006). They faced multiple losses including loss of homes and jobs and loss of or separation from family members and friends. An important lesson learned from these data is that in planning disaster responses, it is crucial for the well-being of children to be mindful of the psychological needs of their parents and to provide as much ongoing support as possible for families as communities work to rebuild.

Results of this study also revealed that children's worries about school were significantly associated with long-term distress. This is consistent with literature on multisystemic pathways associated with trauma recovery (Compas et al., 1995; Sroufe et al., 2005). A recent study related to Hurricane Katrina stress provides support for the relation between school problems and recovery. When eighth-grade students with both test anxiety and PTSD symptoms participated in an intervention that addressed test anxiety, both test anxiety and PTSD symptoms decreased (Weems et al., 2007). Given that multiple systems (e.g., family and school) are affected by trauma, treatment that includes family members and addresses family, school, and peer issues as well as the actual traumatic event may be the most effective intervention and treatment.

Limitations

One of the limitations of the current study, and disaster studies in general, is the lack of predisaster baseline data. Although mental health symptoms were elevated with different outcome trajectories related to the stressors, it is not clear what the premorbid levels of symptoms and functioning were for children and adolescents in the disaster area. Another inherent flaw with longitudinal studies of disaster outcomes is the possibility of incomplete history; that is, some other phenomenon may have caused the symptoms or patterns. Although the lack of statistical significance for pre- and posthurricane losses or traumas in the overall regression model aids in addressing this issue, other factors not measured in this study may affect the outcomes.

Another limitation to this study is generalizability related both to the location of the sample in a relatively small, homogeneous parish with multigenerational family involvement and the severity of the devastation. Without the extent of the devastation, slow recovery, and family and community disruptions in support, it is possible that characteristics of recovery response would be somewhat different. This study also does not account for, and the findings cannot be generalized to, those individuals who did not return or returned but did not remain in the parish. This group is an important population for future research. Also given that fewer students were in the 9–11 age grouping, it may be that families with younger children did not return as rapidly. This may not only be a limitation to the sample but may also account for the lack of interaction between age and gender.

A final limitation is related to the measurement of exposure variables and psychological status. The measurement tool utilized in this survey was the NCTSN Hurricane Assessment and Referral Tool for Children and Adolescents (NCTSN, 2005), a measure based on the validated UCLA PTSD Index that also includes a number of questions regarding depressive symptoms. The overall measure has not been standardized, and although it is theoretically and clinically informed, no psychometric data are available for symptom and cutoff scores. This assessment tool is currently being evaluated.

Future Research

Research on the short- and long-term effects of disasters on children and adolescents, including data from this study related to disaster impact 2–3 years post-Hurricane Katrina, lead to a strong recommendation for additional research related to the psychological effects of disasters on children. Such research needs to include studies of children of different ages, outcome trajectories, and developmental progression. It is also important to continue following the children of Katrina in order to assess longer-term effects and assessments to learn more about students who move from positive to negative postdisaster responses. Such research will advance the science of disaster research and also provide crucial information about ways to better prepare for future disasters and the aftermath. Given the availability of the current data, future research studies will be able to include pre–post data that can then be used to develop meaningful postdisaster interventions and influence policy related to disaster response. This information will also provide more understanding of the effects of different types of disasters, specific response and recovery patterns that can be expected, and the impact on developmental trajectories. The results found in this study provide a step toward looking at the full array of recovery patterns.

This study also highlights the need for more research on factors contributing to resilience. Particular to this study, students who were older, male, did not actively seek mental health treatment, or endorse family and school problems tended to show patterns of positive recovery associated with resilience. Further, students who suffered significant losses following Hurricane Katrina and who had higher levels of storm-related exposure (i.e., injury to self, family, or friends) were more likely to be in the breakdown without recovery or the normal response and recovery group compared with children and adolescents in the stress-resistant group. Future research should address other factors contributing to resilience such as self-efficacy, positive supports, and better coping skills along with exposure to disasters or trauma as it relates to patterns of recovery.

Another promising area of research relates to the evaluation of the effectiveness of developmentally grounded predisaster preparedness on outcomes for children and families (Osofsky et al., 2007). In this parish that was heavily impacted by Hurricane Katrina, patterns of recovery were assessed on several levels. More data are needed in intersecting areas including predisaster functioning, environmental risk, previous exposure to trauma and loss, family disruption, and financial difficulties and their impact on the experience of evacuation, displacement, and subsequent resilience and recovery trajectories. However, considerable information is already available based on findings from this and other studies that will facilitate future planning and preparation in ‘‘disaster-prone’’ areas and developmentally informed responses in preparing for and responding to disasters.

Lessons Learned

It is evident in St. Bernard Parish that the school environment provides a community for children where they not only learn academic skills but also develop social relationships with peers and teachers. For the participants in this study, the positive features of their school environment likely contributed to the resilience and recovery of children despite the ongoing struggles. In spite of the devastation, St. Bernard Parish Unified Schools reopened within 2½ months following Hurricane Katrina with a commitment to being a significant source of support for children and families following the devastation of their community. In disaster planning for children, it is crucial not only to support parents and caregivers but also to focus on building support through schools and community activities. When children returned to school, they were surrounded by peers who had experienced similar trauma; many of those children, particularly those who showed resilience, were likely able to maintain and build their peer groups in this setting following the disaster.

Placing more focus on resilience and recovery in addition to risk and risk mitigation will significantly benefit future research and disaster planning. Given the extent of the devastation and the complexity of recovery resulting from Hurricane Katrina and similar large-scale disasters, we have the opportunity to learn more about these patterns of recovery as well as continuing symptomatology in children, adolescents, and families. Such information will be helpful in planning ways to provide support and enhance development during the recovery period as communities rebuild. Building on these ideas, in a collaborative meeting with the schools, one of the authors of the current study (B. Lawrason, February 2008) noted, "It is not whether our children will recover. It is how they recover, what scarring remains, and how we can help them achieve their potential."

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012–1028.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: American Psychiatric Association.

Bal, A. (2008). Post-traumatic stress disorder in Turkish child and adolescent survivors three years after the Marmara Earthquake. *Child and Adolescent Mental Health*, *13*, 134–139.

Blos, P. (1967). The second individuation process of adolescence. *Psychoanalytic Study of the Child*, *22*, 162–186.

Bokszczanin, A. (2007). PTSD symptoms in children and adolescents 28 months after a flood: Age and gender differences. *Journal of Traumatic Stress*, *20*, 347–351.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist*, *59*, 20–28.

Bonanno, G. A., & Mancini, A. D. (2008). The human capacity to thrive in the face of potential trauma. *Pediatrics*, *121*, 369–375.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychopathology*, *22*, 723–742.

Compas, B. E., Hinden, B. R., & Gerhardt, C. A. (1995). Adolescent development: Pathways and processes of risk and resilience. *Annual Review of Psychology*, *46*, 265–293.

Cryder, C. H., Kilmer, R. P., Tedeschi, R. G., & Calhoun, L. G. (2006). An exploratory study of posttraumatic growth in children following a natural disaster. *American Journal of Orthopsychiatry*, *76*, 65–69.

Garrison, C. Z., Weinrich, M. W., Hardin, S. B., Weinrich, S., & Wang, L. (1993). Post-traumatic stress disorder in adolescents after a hurricane. *American Journal of Epidemiology*, *138*, 522–530.

Giannopoulou, I., Strouthos, M., Smith, P., Dikaiakou, A., Galanopoulou, V., & Yule, W. (2006). Post-traumatic stress reactions of children and adolescents exposed to the Athens 1999 earthquake. *European Psychiatry*, *21*, 160–166.

Goenjian, A. K., Molina, L., Steinberg, A. M., Fairbanks, L. A., Alvarez, M. L., Goenjian, H. A., et al. (2001). Posttraumatic stress and depressive reactions among Nicaraguan adolescents after Hurricane Mitch. *American Journal of Psychiatry*, *158*, 788–794.

Goenjian, A. K., Walling, D., Steinberg, A. M., Karayan, I., Najarian, L. M., & Pynoos, R. (2005). A prospective study of posttraumatic stress and depressive reactions among treated and untreated adolescents 5 years after a catastrophic disaster. *American Journal of Psychiatry*, *162*, 2302–2308.

Green, B. L., Grace, M. C., Vary, M. G., Kramer, T. L., Gleser, G. C., & Leonard, A. C. (1994). Children of disaster in the second decade: A 17-year follow up of Buffalo Creek survivors. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 71–79.

Halpern, J., & Tramontin, M. (2007). *Disaster mental health: Theory and practice*. Belmont, CA: Brooks/Cole.

Hamada, R. S., Kameoka, V., Yanagida, E., & Chemtob, C. M. (2003). Assessment of elementary school children for disaster-related posttraumatic stress disorder symptoms: The Kauai recovery index. *Journal of Nervous and Mental Disease*, *191*, 268–272.

John, P. B., Russell, S., & Russell, P. S. (2007). The prevalence of posttraumatic stress disorder among children and adolescents affected by tsunami disaster in Tamil Nadu. *Disaster Management and Response., 5*, 3–7.

Kar, N., & Bastia, B. K. (2006). Post-traumatic stress disorder, depression and generalized anxiety disorder in adolescents after a natural disaster: A study of comorbidity. *Clinical Practice and Epidemiology in Mental Health, 2*. Retrieved May 11, 2010, from http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1563462&blobtype=pdf

Kar, N., Mohapatra, P. K., Nayak, K. C., Pattanaik, P., Swain, S. P., & Kar, H. C. (2007). Post- traumatic stress disorder in children and adolescents one year after a super-cyclone in Orissa, India: Exploring cross-cultural validity and vulnerability factors. *BMC Psychiatry, 14*, 7–8.

Kessler, R. C., Galea, S., Jones, R. T., & Parker, H. A. (2006). Mental illness and suicidality after Hurricane Katrina. *Bulletin of the World Health Organization, 84*, 930–939.

Knabb, R. D., Rhome, J. R., & Brown, D. P. (2006). *Tropical cyclone report Hurricane Katrina*. National Hurricane Center. Retrieved May 11, 2010, from http://www.nhc.noaa.gov/pdf/TCR-AL122005_Katrina.pdf

Kolaitis, G., Kotsopoulos, J., Tsiantis, J., Haritaki, S., Rigizou, F., Zacharaki, L., et al. (2003). Posttraumatic stress reactions among children following the Athens earthquake of September 1999. *European Child and Adolescent Psychiatry, 12*, 273–280.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Prinstein, M. J. (1996). Symptoms of posttraumatic stress in children after Hurricane Andrew: A prospective study. *Journal of Consulting and Clinical Psychology, 64*, 712–723.

Lonigan, C. J., Shannon, M. P., Taylor, C. M., Finch, A. J., & Sallee, F. R. (1994). Children exposed to disaster II: Risk factors for the development of post-traumatic symptomatology. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*, 94–105.

Marsee, M. A. (2008). Reactive aggression and posttraumatic stress in adolescents affected by Hurricane Katrina. *Journal of Clinical Child and Adolescent Psychology, 37*, 519–529.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*, 227–238.

Masten, A. S., & Obradovic, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society, 13*. Retrieved May 11, 2010, from http://www.ecologyandsociety.org/vol13/iss1/art9/

McDermott, B. M., & Palmer, L. J. (2002). Postdisaster emotional distress, depression and event- related variables: Findings across child and adolescent developmental stages. *Australian and New Zealand Journal of Psychiatry, 36*, 754–761.

McFarlane, A. C., Policansky, S. K., & Irwin, C. (1987). A longitudinal study of the psychological morbidity in children due to a natural disaster. *Psychological Medicine, 17*, 727–738.

National Child Traumatic Stress Network. (2005). *Hurricane Assessment and Referral Tool for Children and Adolescents*. Retrieved May 11, 2010, from http://www.nctsnet.org/nctsn_assets/pdfs/intervention_manuals/referraltool.pdf

Neuner, F., Schauer, E., Catani, C., Ruf, M., & Elbert, T. (2006). Post-tsunami stress: A study of posttraumatic stress disorder in children living in three severely affected regions in Sri Lanka. *Journal of Traumatic Stress, 19*, 339–347.

O'Daugherty, M., & Masten, A. S. (2005). Resilience processes in development: Fostering positive adaptation in the context of adversity. In S. Goldstein & R. B. Brooks (Eds.), *Handbook of resilience in children* (pp. 17–38). New York: Springer.

Osofsky, H. J., Osofsky, J. D., Kronenberg, M., Brennan, A., & Hansel, T. C. (2009). Posttraumatic stress symptoms in Children after Hurricane Katrina: Predicting the need for mental health services. *American Journal of Orthopsychiatry, 79*, 212–220.

Osofsky, J. D. (Ed.). (2004). *Young children and trauma: Intervention and treatment*. New York: Guilford.

Osofsky, J. D., Osofsky, H. J., & Harris, W. W. (2007). Katrina's children: Social policy considerations for children in disasters. *Social Policy Report, 21*, 3–18.

Osofsky, J. D., Osofsky, H. J., Kronenberg, M., & Cross, T. (2009). The aftermath of Hurricane Katrina: Mental health considerations and lessons learned. In R. P. Kilmer, V. Gil-Rivas, R. G. Tedeschi, & L. G. Calhoun (Eds.), *Meeting the needs of children, families, and communities post-disaster: Lessons learned from Hurricane Katrina and its aftermath* (pp. 241–263). Washington, DC: American Psychological Association.

Pfefferbaum, B. (1997). Posttraumatic stress disorder in children: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1503–1511.

Piyasil, V., Ketuman, P., Plubrukarn, R., Jotipanut, V., Tanprasert, S., Aowjinda, S., et al. (2007). Post traumatic stress disorder in children after tsunami disaster in Thailand: 2 years follow-up. *Journal of the Medical Association of Thailand, 90*, 2370–2376.

Pynoos, R. S. (1993). Traumatic stress and developmental psychopathology in children and adolescents. In J. M. Oldham, M. B. Riba, & A. Tasman (Eds.), *Review of psychiatry* (pp. 205–238). Washington, DC: American Psychiatric Publishing.

Pynoos, R. S., Goenjian, A., Tashjian, M., Karakashian, M., Manjikian, R., Manoukian, G., et al. (1993). Posttraumatic stress reactions in children after the 1988 Armenian earthquake. *The British Journal of Psychiatry, 163*, 239–247.

Pynoos, R. S., Steinberg, A. M., & Piacentini, J. C. (1999). A developmental psychopathology model of childhood traumatic stress and intersection with anxiety disorders. *Biological Psychiatry, 46,* 1542–1554.

Roussos, A., Goenjian, A. K., Steinberg, A. M., Sotiropoulou, C., Kakaki, M., Kabakos, C., et al. (2005). Posttraumatic stress and depressive reactions among children and adolescents after the 1999 earthquake in Ano Liosia, Greece. *American Journal of Psychiatry, 162,* 530–537.

Russoniello, C. V., Skalko, T. K., O'Brien, K., McGhee, S. A., Bingham-Alexander, D., & Beatley, J. (2002). Childhood posttraumatic stress disorder and efforts to cope after Hurricane Floyd. *Behavioral Medicine, 28,* 61–71.

Scheeringa, M. S., & Zeanah, C. H. (2008). Reconsideration of harm's way: Onsets and comorbidity patterns of disorders in preschool children and their caregivers following Hurricane Katrina. *Journal of Clinical Child and Adolescent Psychology, 37,* 508–518.

Shannon, M. P., Lonigan, C. J., Finch, A. J., & Taylor, C. M. (1994). Children exposed to disaster: I: Epidemiology of post-traumatic symptoms and symptom profiles. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 80–93.

Shaw, J. A. (2000). Children, adolescents, and trauma. *Psychiatric Quarterly, 71,* 227–243.

Spell, A. W., Kelley, M. L., Wang, J., Self-Brown, S., Davidson, K., Pellegrin, A., et al. (2008). The moderating effects of maternal psychopathology on children's adjustment post-Hurricane Katrina. *Journal of Clinical Child and Adolescent Psychology, 37,* 553–563.

Sroufe, L. A. (1997). Psychopathology as an outcome of development. *Development and Psychopathology, 9,* 251–268.

Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2005). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood.* New York: Guilford.

St. Bernard Parish Government. (2008). *St. Bernard issues mid-year State of the Parish: Charting our recovery 2008.* Retrieved May 11, 2010, from http://www.sbpg.net/index.php?option=com_content&view=article&id=206:mid-year-report&catid=2:local-news&Itemid=2

St. Bernard Parish Public Schools. (2008). *The rebuilding of St. Bernard Parish Public Schools.* Retrieved May 11, 2010, from http://www.stbernard.k12.la.us/downloads/0809_RebuildingPlan.pdf

Steinberg, A. M., Brymer, M. J., Decker, K. B., & Pynoos, R. S. (2004). The University of California at Los Angeles Post-Traumatic Stress Disorder Reaction Index. *Current Psychiatry Reports, 6,* 96–100.

Terranova, A. M., Boxer, P., & Morris, A. S. (2009). Factors influencing the course of posttraumatic stress following a natural disaster: Children's reactions to Hurricane Katrina. *Journal of Applied Developmental Psychology, 30,* 344–355.

Thienkrua, W., Cardozo, B. L., Chakkraband, M. L., Guadamuz, T. E., Pengjuntr, W., Tantipiwatanaskul, P., et al. (2006). Symptoms of posttraumatic stress disorder and depression among children in tsunami-affected areas in southern Thailand. *Journal of the American Medical Association, 296,* 549–559.

Vogel, J. M., & Vernberg, E. M. (1993). Children's psychological responses to disaster. *Journal of Clinical Child Psychology, 22,* 464–484.

Weems, C. F., Pina, A. A., Costa, N. M., Watts, S. E., Taylor, L. K., & Cannon, M. F. (2007). Predisaster trait anxiety and negative affect predict postraumatic stress in youths following Hurricane Katrina. *Journal of Consulting and Clinical Psychology, 75,* 154–159.

Weems, C. F., Taylor, L. K., Costa, N. M., Marks, A. B., Romano, D. M., Verrett, S. L., et al. (2009). Effect of school-based test anxiety intervention in ethnic minority youth exposed to Hurricane Katrina. *Applied Developmental Psychology, 30,* 218–226.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1260–1270

Children With Disabilities in the Context of Disaster: A Social Vulnerability Perspective

Lori Peek
Colorado State University

Laura M. Stough
Texas A&M University

An estimated 200 million children worldwide experience various forms of disability. This critical review extrapolates from existing literature in 2 distinct areas of scholarship: one on individuals with disabilities in disaster, and the other on children in disaster. The extant literature suggests that various factors may contribute to the physical, psychological, and educational vulnerability of children with disabilities in disaster, including higher poverty rates, elevated risk exposure, greater vulnerability to traumatic loss or separation from caregivers, more strain on parents, and poor postdisaster outcomes, unless medical, familial, social, and educational protections are in place and vital social networks are quickly reestablished. Future research needs are outlined in the conclusion.

Research on children and disaster has increasingly examined how certain characteristics, such as age and developmental levels, gender, and ethnicity, intersect and affect children's experiences in disaster (La Greca, Silverman, Vernberg, & Roberts, 2002; Peek, 2008). However, researchers have rarely considered how disability may contribute to the short- and long-term impacts of disaster on children. This represents a significant gap in knowledge, given that in the United States almost 9.2% of the school-aged population receives special education services (U.S. Department of Education, 2009), and an even higher percentage of children with disabilities are estimated to live in developing countries (World Health Organization [WHO], 2005).

In this review of the existing literature, we examine the special situation of children with disabilities in the context of disaster, highlighting research on the potential vulnerabilities, risks, and resiliency factors of this population with the goal of building a better knowledge base for future research in this area. We begin with an overview of the social vulnerability perspective and then summarize definitions of disability that are often used by scholars and practitioners working on disability issues in disaster. The subsequent sections of the article draw on empirical studies from disaster researchers and focus on the adverse physical, psychological, and educational vulnerabilities that children with disabilities may experience in the context of disaster. We conclude with a discussion of the theory of resilience as it pertains to children with disabilities in disaster and outline additional research needs.

Social Vulnerability and Disaster

Historically, disasters have been viewed as random events that cause indiscriminate losses and suffering among affected populations. However, in the mid-1970s, social scientists who pioneered the social vulnerability perspective began to question the ''naturalness'' of ''natural disasters'' (O'Keefe, Westgate, & Wisner, 1976). Proponents of the social vulnerability school assert that disasters are in fact a complex mix of natural hazards and human action (Cutter, Boruff, & Shirley, 2003; Wisner, Blaikie, Cannon, & Davis, 2004). Disasters are thus conceptualized as the product of the convergence of social, political, and economic factors that shape people's exposure to risk and ability to prepare for,

---

The authors are listed alphabetically to denote equal contributions.

This work was supported by the National Institute of Mental Health Research Education in Disaster Mental Health (REDMH) mentoring program, Grant 5R25MH068298-04. The authors would like to thank Fran Norris, Carol North, and Larry Palinkas for their guidance.

Correspondence concerning this article should be addressed to Lori Peek, Department of Sociology, Colorado State University, B-237 Clark Building, Fort Collins, CO 80523-1784. Electronic mail may be sent to lori.peek@colostate.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0016

respond to, and recover from extreme events (Wisner et al., 2004, p. 4). This theoretical frame posits that we as a human society, rather than nature alone, decide who is more likely to live near or in the path of dangerous agents and to have weakened or no defenses against them (Hewitt, 1997, p. 141).

Disaster risk is socially distributed in ways that reflect preexisting inequalities, in that some groups are more prone to death, injury, economic loss, and psychological impairment in the wake of differing hazards (Wisner et al., 2004). Specifically, children, the elderly, women, racial minorities, the poor, persons with physical or mental disabilities, and immigrants have been identified by both disaster researchers and policy makers as especially vulnerable to the harmful impacts of disaster (Cutter et al., 2003). Although these groups differ in many ways, when viewed through a social vulnerability perspective they all can be seen as lacking access to vital economic and social resources, possessing limited autonomy and power, and having low levels of social capital (Morrow, 1999). They also, for sociohistorical and economic reasons, tend to live in the most hazardous regions and in the lowest quality housing, further exposing them to risks associated with natural hazards (Cutter et al., 2003; Wisner et al., 2004). Due to the disproportionate losses members of these marginalized groups suffer in disaster, they consequently have the hardest time recovering, creating vicious cycles of loss and vulnerability (United Nations Human Settlements Programme, 2007).

The social vulnerability paradigm, and the ensuing identification of vulnerable groups, has helped focus attention on those most at risk during and in the aftermath of disaster. Yet, social science research on vulnerable populations in disaster has tended to concentrate on single demographic factors—such as age, race, gender, or socioeconomic status—instead of examining the complex intersections between these categories (Phillips & Morrow, 2007). For example, it is clear that not all children are equally vulnerable to the impacts of disaster. Instead, the age of the child intersects with other personal and social characteristics, such as his or her geographic location, family structure, socioeconomic status, physical and mental ability, stage of development, and nationality, to determine the likelihood of harm in a particular disaster event. Given that these vulnerability factors tend to "cluster," certain segments of the population, such as children with disabilities, may experience amplified risk in disaster.

Defining Disability

Disability is a broad term that is not consistently defined. Its use varies according to societal norms and by the organizations, medical entities, and governmental agencies that address disability issues. The limited extant research on the effects of disaster on individuals with disabilities similarly has defined disability in a variety of ways. Mental health researchers typically use criteria from the American Psychiatric Association's (2000) *Diagnostic and Statistical Manual* to define samples with cognitive and psychological disabilities. Disaster researchers who study individuals with mobility impairments tend to use the Americans With Disabilities Act (1990) definition of disability as being "a physical or mental impairment that substantially limits one or more of the major life activities of such individual" (PL 101-336, 104 Stat. 327). Given their focus on disaster population statistics, epidemiologists often rely on the U.S. Census Bureau definition of disabilities, which allows for the analysis of data on persons with sensory, physical, mental, self-care, mobility, and employment-related disabilities. The emergency management field categorizes people who use wheelchairs, non-English speakers, and the elderly together as "special needs" populations.

Considering these varying taxonomies, definitions of disability in disaster research are best subsumed under the WHO's International Classification of Functioning, Disability, and Health (WHO, 2001). The ICF conceptualizes disability as resulting from the interaction between the health condition of an individual and that individual's personal and environmental setting. Thus defined, disability refers to an impairment in body functions or structures, a limitation in a specific activity, or a restriction in social participation. To date, research on the effects of disaster on people with disabilities has almost exclusively focused on how physical or cognitive impairments intersect with disaster experiences rather than on the interplay of disability with social and environmental factors. Kailes and Enders (2007) have proposed a functional-needs approach to defining disability in disaster, which highlights the supports individuals may need in the areas of communication, medical health, functional independence, supervision, and transportation. This functional-needs approach to disability was adopted by the U.S. Federal Emergency Management Agency and Department of Homeland Security Office for Civil Rights and Civil Liberties (2008) as part of its Comprehensive Preparedness

Guide 301 (CPG-301), which is designed to aid tribal, territorial, state, and local governments in planning for individuals with special needs during an emergency. The functional-needs approach has not yet been used, however, as a theoretical approach in disaster research on individuals with disabilities.

Identifying the precise number of individuals with disabilities within a particular sociocultural context also depends on the definition chosen. According to the WHO (2005), roughly 600 million people—10% of the global population—have some type of disability. Disability is highly correlated with poverty, and as many as 80% of all individuals with disabilities live in developing countries. The U.S. Census Bureau estimates the prevalence of disability in children between the ages of 6 and 15 to be nearly 13% (U.S. Census Bureau, 2005). The proportion of children with disabilities in developing countries is likely significantly higher, and it is estimated that over 200 million children worldwide have some type of disability (United Nations International Children's Emergency Fund, 2007).

Children, Disability, and Disaster

By the end of the 20th century, disasters affected an estimated 67 million children around the world each year (Penrose & Takaki, 2006). This number is projected to more than triple over coming decades, mostly due to population growth, especially in hazardous regions such as low-lying coastal zones, and the growing prevalence and severity of climate change-related disasters (Bartlett, 2008). Conservative estimates suggest that over 7 million of these disaster-affected children experience various forms of disability (United Nations International Children's Emergency Fund, 2007), and millions more may acquire disabilities as a result of increasingly frequent and intense disaster activity, wars, and landmine explosions (World Health Organization, 2005).

Scholarship on social vulnerability and disaster often lists both children and individuals with disabilities as populations at particular risk when disaster strikes. Consequently, two parallel tracks of research—one on children, the other on people with disabilities—have emerged within the disaster literature. Although an increasing amount of scholarly and applied attention has focused on the implications of disaster for children and youth, little empirical work has examined the effects of disaster on individuals with disabilities. The research that is available has focused on adult populations (Stough, 2009) and seldom mentions the needs of children with disabilities.

Given the dearth of research that reports on children with disabilities in disaster, we extrapolate from work in both areas to examine the potential implications of disasters for children with disabilities. To frame our discussion of the vulnerability and risk factors for children with disabilities, we draw on Peek's (2008) model that has previously been used to review research on children in disasters in terms of their physical, psychological, and educational vulnerabilities.

Physical Vulnerability

When disaster strikes, children are among those most vulnerable to death and injury, especially in developing countries (Neumayer & Plümper, 2007). Similarly, adults with disabilities or mobility impairments are at increased risk for mortality (Aldrich & Benson, 2008). Children and individuals with disabilities share a number of common risk factors that increase the probability that they will experience negative physical impacts during and following disaster.

First, children and people with disabilities are more likely to live in poverty both in developed and developing nations. In the United States, a higher percentage of children (18.3%) and individuals with severe disabilities (26%) live in poverty than the population as a whole (13.3%; U.S. Census Bureau, 2006). In most high-income countries, people under 18 years of age make up about 20% of the population; in low-income countries, where children are more likely to experience disability, children represent closer to half of the total population (e.g., 42% in Bangladesh, 51% in Nigeria, 57% in Uganda; Bartlett, 2008, p. 75). Poor people around the world are more physically vulnerable to natural disasters because they cannot afford to engage in prescribed preparedness actions, often do not receive or understand warning messages, are less likely to have the resources to evacuate a threatened area, and typically live in low-cost, lower-quality housing that is more prone to damage or collapse (Fothergill & Peek, 2004; Wisner et al., 2004). Children with disabilities in high-poverty areas are thus among those most exposed and most vulnerable to extreme events.

Second, in the case of sudden onset disasters that allow little warning time, such as tornados or earthquakes, children with disabilities may have a more difficult time taking recommended protective actions, escaping, or withstanding the force of the disaster. For instance, children with mobility limitations may be incapable of crouching under their desks in an earthquake, hiking up a hillside in a flash flood, or running to an evacuation point on higher ground in the event of a tsunami. All of these are recommended protective actions developed for people without mobility limitations. Children with cognitive impairments may not recognize signs of environmental danger or understand impending threats (Kailes & Enders, 2007) or may become anxious and confused in response to emergency signals (Scotti et al., 2007).

Third, even when evacuation is possible due to longer warning periods, such as with hurricanes, children with disabilities may be less likely to leave the threatened area thus putting them at increased risk for death or injury. Although much prior disaster research has shown that adults with children are *more likely* to respond to disaster warnings and evacuation orders than people without children, research on adults with disabilities has revealed that they are among those *least likely* to evacuate (see Dash & Gladwin, 2007). For example, studies on adults with sensory disabilities such as blindness or deafness suggest that these individuals often do not receive timely warning messages (Phillips & Morrow, 2007) and that they encounter significant barriers during evacuation (White, 2006). Van Willigen, Edwards, Edwards, and Hessee (2002) contrasted the experiences of households that had a family member with a disability with other households in North Carolina in the aftermath of Hurricanes Bonnie, Dennis, and Floyd. They found that evacuation rates were anywhere from 9% to 25% lower among those households that had a member of the family with a disability. These households were more likely to report that transportation issues and the lack of accessible shelters contributed to their decision not to evacuate. Research has also revealed that people with disabilities are less apt to have evacuation plans than those without disabilities (Spence, Lachlan, Burke, & Seeger, 2007).

Fourth, both children and people with disabilities are often excluded from emergency preparedness planning at all levels of government, leaving children with disabilities especially unprepared for emergencies. Disaster response professionals commonly assume that parents will inform, warn, and protect children in the event of disaster, even though children are frequently apart from their parents when in school, daycare, or with their friends (Mitchell, Haynes, Hall, Choong, & Oven, 2008; Phillips & Morrow, 2007). Likewise, emergency management agencies regularly do not include the needs of individuals with disabilities in their emergency preparedness policies, citing factors such as inadequate personnel or their own limited knowledge of disability (Fox, White, Rooney, & Rowland, 2007; Rowland, White, Fox, & Rooney, 2007). This lack of attention to disability-related needs in emergency management could result in children with disabilities being left behind in an evacuation or forced to evacuate without vital supports (e.g., mobility devices, respirators, medications, companion animals, etc.; Osofsky, Osofsky, & Harris, 2007; Rooney & White, 2007). As a case in point, approximately 100,000 children and their families evacuated from the city of New Orleans as a result of Hurricane Katrina (Casserly, 2006). Over one third of the children who remained displaced months after the storm had at least one diagnosed chronic medical condition, and their access to health insurance, continuous medical care, prescription medication, and specialized medical equipment was significantly compromised (Abramson & Garfield, 2006; Abramson, Redlener, Stehling-Ariza, & Fuller, 2007).

Fifth, children with disabilities may be in danger of suffering life-threatening consequences in the aftermath of disaster due to separation from parents and other caregivers. Children who are separated from their parents after disaster are more prone to illness and disease, malnutrition, and abuse, and may have difficulty in situations with prolonged deprivation associated with more chronic events such as droughts (Babugura, 2008; Bartlett, 2008). Individuals with physical disabilities, which include those with paralysis or who are medically fragile, are similarly susceptible to secondary illnesses in conditions in which their ongoing health needs are not appropriately addressed (Kinne, Patrick, & Doyle, 2004).

Sixth, the stigma, social distancing, and institutional exclusion often experienced by individuals with disabilities can further threaten the physical health and well-being of children with disabilities in the aftermath of disaster (Hemingway & Priestly, 2006; Tierney, Petak, & Hahn, 1988). For example, children with disabilities may acquire additional impairments and experience health issues as a result of inadequately staffed shelters that are not prepared to meet their medical needs. Indeed, the National Organization on Disability (2005) reports that shelters often do not offer equitable access to services for people with disabilities, which may also mean this group is overlooked in the distribution of basic relief and excluded from full participation in response and recovery activities (Hemingway & Priestly, 2006). In some developing countries, individuals with disabilities are socially stigmatized and face considerable social discrimination (Institute of Medicine, 2001), and anecdotal accounts suggest that some individuals with disabilities may even be purposely abandoned during disaster. For instance, news media reported that people with physical and cognitive disabilities were left behind during evacuation in the 2000 floods in Mozambique and Zimbabwe (United Nations Human Settlements Programme, 2007).

Psychological Vulnerability

Most of the research that has been conducted on children and disasters has explored their psychological vulnerability in the short-term aftermath of disaster (La Greca et al., 2002; Weissbecker, Sephton, Martin, & Simpson, 2008). This body of work collectively indicates that school-aged children are often more psychologically affected by disaster than adults and that children of different ages express trauma differently. Following direct exposure to disaster, young children may exhibit increases in aggressive and oppositional behaviors, separation anxiety, diminished activity levels, somatic complaints, and changes in sleeping behaviors, whereas adolescents tend to have elevated anxiety levels, diminished academic performance, and may engage in delinquent behaviors (Norris et al., 2002). It should be noted, however, that most children, even when directly impacted by disaster, do *not* develop new psychopathologies. Children can and do effectively adapt to disruptive shocks caused by disaster, especially when they receive appropriate support from a parent or other close adult and live in a safe and nurturing environment (Dubow & Tisak, 1989; La Greca et al., 2002; Masten, 2001).

Very little research has been conducted on the psychological effects of disaster on individuals with disabilities. Some policy accounts suggest that adults with disabilities experience feelings of distress, depression, or anxiety following disaster (Rooney & White, 2007), but most of these reports did not include a formal psychological assessment. However, a few studies on individuals with preexisting mental illnesses have found that when psychological treatment is continued postdisaster, patients do not acquire new psychopathologies, nor do their symptoms intensify (Bystritsky, Vapnik, Maidment, Pynoos, & Steinberg, 2000; McMillen, North, Mosley, & Smith, 2002).

Most children with disabilities are not only reliant on their parents for the same types of care as their peers who do not have disabilities but also often depend on their parents for additional functional and emotional supports. For instance, children with health care needs often have their parents administer prescriptions or perform medically related functions such as tube feeding or transferring them in and out of a wheelchair. Children with sensory disabilities such as blindness or deafness may rely on their parents for communication needs, and those with intellectual disabilities may need their parents to explain unfamiliar events when they occur. Yet trauma can have a deleterious impact on mothers and fathers and may limit their capacity to parent their children effectively (Appleyard & Osofsky, 2003). When parents suffer from depression, anxiety, posttraumatic stress disorder (PTSD), or other postdisaster disturbances, they are more likely to feel helpless and frustrated as they attempt to meet the support needs of their children (Osofsky, 1995), and severe parental distress can, in turn, contribute to PTSD symptomatology in children (Breton, Valla, & Lambert, 1993; Green et al., 1991). Given the added demands on parents of children with disabilities, it is possible that both these adults and their children could be at an even higher risk for postdisaster psychological impairment.

The separation of a parent and child during a disaster can be especially distressing to both the adult and child. For instance, one post-Katrina study found that trauma symptoms consistent with depression and PTSD were higher among children separated from caregivers during evacuation and displacement than among those children whose families remained intact (Osofsky et al., 2007). Because children with disabilities have social networks that tend to be smaller and more fragile (Smart, 2001), it is understandable that losing a parent or other person in their limited support system due to physical separation or even death could be particularly traumatic.

Research on individuals with disabilities suggests that they and their family members have more stressful evacuation and sheltering experiences and can be at a disadvantage in understanding instructions and procedures in these situations (Barile, Fichten, Ferraro, & Judd, 2006; Peek & Fothergill, 2008; White, 2006). Scotti et al. (2007) found that emergency situations, which typically include startling visual and auditory alerting systems and involve the presence of emergency personnel and other strangers, trigger negative stress reactions in adults with intellectual disabilities. Children with intellectual disabilities such as autism may become agitated by alert systems and postdisaster sheltering. Children are often excluded in terms of postdisaster communication and decision making (Mitchell et al., 2008), and children with disabilities may be especially prone to exclusion from information and services made available to other children in shelters such as recreation, crisis intervention, or different forms of therapy.

Due to their economic status, individuals with disabilities are more likely to live in areas that are at risk of physical impact during disaster (Hewitt, 2007; Wisner et al., 2004), and therefore it is probable that children with disabilities and their families will experience significant material loss. Proximity to a disaster and direct exposure to destruction, trauma, and loss, in turn, increases the likelihood that a child will develop PTSD (Garfinkel, Kaushal, Teitler, & Garcia, 2005; Norris et al., 2002; Weissbecker et al., 2008). In their posthurricane research, Van Willigen et al. (2002) found that a higher percentage of households that included an individual with a disability experienced damage to their homes, and that members of these households were more likely to report that their lives were still disrupted more than a year after the disaster. Given children's elevated risk for trauma following disaster, children with disabilities who are directly exposed to disaster may be more susceptible to long-term psychological disruptions and they may exhibit negative reactions years following the event (Christ & Christ, 2006). In addition, although research has yet to explore whether children with disabilities are at risk for experiencing recurring disasters, the compounded vulnerability of this population suggests that this may be the case.

Although research shows that most children can and do recover psychologically from disaster, children with disabilities may face additional barriers that may delay their recovery. In their analysis of survey data with New York City residents, Garfinkel et al. (2005) found that adults with work disabilities and children with activity limitations were among those most likely to report persistent mental and physical health problems 6 months after the 9/11 attacks. According to Stough and Sharp (2007), adults with disabilities are more likely to require accessible housing and social services postdisaster to support their disability-related needs as well as a substantially longer period of case management than do adults without disabilities. Following Hurricane Katrina, children with cognitive disabilities who were displaced from their home communities were placed on long waiting lists for psychiatric care. In addition, several parents noted that it was increasingly difficult or impossible for them to maintain their children's prescribed psychotropic medications, either because they could not find and afford appropriate psychiatric help or their medical records had been lost (Abramson & Garfield, 2006).

In a review of studies of child psychiatric interventions following disaster, La Greca (2001) concludes that when children are able to cognitively reprocess their traumatic experience, treatment of severe levels of PTSD is more successful. A challenge for children with cognitive disabilities, such as mental retardation or autism, is that they have difficulties in processing information, which may impede their psychological recovery postdisaster. Tierney et al. (1988) suggest that the social distancing associated with the label of "disabled" may further limit access to social networks and others sources of psychological support during a disaster. This same social distancing may also make it difficult for a child with a disability to adjust emotionally to a new school, neighborhood, or community.

Educational Vulnerability

Childhood is a time of rapid intellectual development, and thus children are uniquely vulnerable to negative educational outcomes that may accompany disaster (see Peek, 2008). Our search of the literature revealed limited research on the specific educational impacts of disaster for children with disabilities, likely because almost all of the extant literature on disability and disaster focuses on adult populations. The exception was Christ and Christ's (2006) post-9/11 study of children with learning disabilities. All of the children in their sample were living in one of the five boroughs of New York City at the time of the terrorist attacks and lost their firefighter fathers in the disaster. The researchers found that the children struggled with behavioral issues and increased educational demands after the disaster, although supportive special educators and school environments helped mitigate some of the most severe impacts of their losses.

The literature on the effects of disaster on children's education, in comparison, is much broader. Disasters often destroy school buildings, especially in locations where engineering standards and building codes are not enforced or where buildings are of less structural integrity: Hewitt (2007) inventoried tens of thousands of schools that collapsed in earthquakes over the past two decades in Pakistan, India, Nepal, Egypt, Turkey, Armenia, China, and other countries. The loss of schools may leave surviving children with few alternatives for an adequate education, and this issue may be especially problematic for children with disabilities as they face many barriers in accessing education during nondisaster times in both developed and developing countries (International Federation of Red Cross and Red Crescent Societies, 2007). When disaster disrupts a community's educational system, schools accessible to children with disabilities may be among the last to be rebuilt and ramps and other forms of access into inclusive schools may become a secondary priority.

After a disaster, teachers and other key school personnel who remain in the affected region may become overburdened and preoccupied with loss and increasing personal and professional demands (Osofsky et al., 2007). Educators may also become displaced, choose not to return to the community, or lose their jobs due to budget shortfalls (Casserly, 2006). The loss of experienced teachers and other typical supporters of children can be disruptive to children's academic progress, and may be especially challenging for children with disabilities. These children and their families often have complex support networks that include health care professionals, education specialists, psychologists, and other experts in their school district before disaster strikes. There is a chronic shortage of special education teachers worldwide, including in the United States, and replacing qualified educators in the postdisaster recovery period can be quite challenging. Competing demands for limited resources in disaster-affected school districts could further delay the hiring of these personnel. Students with disabilities can be highly sensitive to discontinuities in their educational services, and prolonged disruptions could ultimately diminish their long-term educational outcomes. Children with disabilities also tend to have challenges in establishing social relationships, and therefore may experience the loss of their educational support network more acutely (Smart, 2001).

Vital school records are often damaged or destroyed in a disaster, which can delay the enrollment of students in new schools (Picou & Marshall, 2007). Beyond their general school records, children with disabilities typically have extensive educational and medical records, as well as Individual Education Plans, that may be lost. The loss of these documents can pose significant challenges for new teachers who rely on diagnostic assessments of performance to recreate adequate educational plans for students with special needs.

Environmentally displaced children in developing countries are sometimes forced to stop attending school in order to assist with household chores or to participate in wage labor (Babugura, 2008). In the United States, children displaced by Hurricane Katrina suffered numerous educational challenges, including multiple school relocations, unstable family situations, and unwelcoming and hostile classroom environments (Fothergill & Peek, 2006; Picou & Marshall, 2007). Counselors and teachers who worked with Katrina-affected youth described worrisome behaviors in both elementary and secondary school students, including an inability to pay attention in class, disruptive actions, changes in their friendships, and increased risk taking (Osofsky et al., 2007). Individuals with disabilities are generally more vulnerable to dislocating events (Garfinkel et al., 2005). As a result, children with disabilities may be at special risk for educational delays, diminished academic achievement, or negative behavioral responses in the classroom.

Vulnerability and Resilience: Future Research Needs

Our overview of the relevant literature in two distinct areas of scholarship—one on individuals with disabilities in disaster, the other on children in disaster—points to a number of factors that may contribute to the physical, psychological, and educational vulnerability of children with disabilities in the context of disaster. Specifically, children with disabilities may be expected to show: higher poverty rates, elevated exposure to hazards, greater vulnerability in the context of traumatic loss or separation from caregivers, more strain on parents, and worse postdisaster outcomes unless special medical, familial, social, and educational protections are in place and vital social networks can be quickly reestablished. Consistent with a bioecological systems approach (e.g., Bronfenbrenner & Ceci, 1994; Cicchetti & Lynch, 1993), these factors are interactive and therefore create a *layering of vulnerabilities* that may place children with disabilities differentially at risk before, during, and in the aftermath of disaster.

This review has highlighted the potential vulnerabilities of children with disabilities in disaster as well as the social–ecological nature of that vulnerability. Due to the limited research on the effects of disaster on children with disabilities, our review has said little about the resilience of children with disabilities following disaster. Masten and

Obradovic (2008) have previously used resilience theory to examine how systems interact with individuals in disaster-related events and to delineate different patterns of individual functioning in the aftermath of disaster. In their conceptualization, resiliency refers to how individuals regain normal functioning following exposure to disaster and is described in terms of "recovery, bouncing back, normalization, or self-righting" (p. 9). When considering recovery postdisaster, we can anticipate that children with disabilities will follow one of Masten and Obradovic's patterns in which individuals enter the disaster impact stage already demonstrating a lower level of adaptive functioning. This suppressed level of functioning may also result from, as we have discussed, the interaction between the preexisting condition of the child with the external risk factors to which they were exposed predisaster. The developmental trajectory that children with disabilities will subsequently follow postdisaster, however, is less obvious.

Existing research illuminates the reliance of children with disabilities upon the broader macrostructure to protect them from disaster and it is likely that they will be equally dependent upon this same macrostructure to help them recover in the aftermath of disaster. Thus, when considering resilience postdisaster, we can predict that without a home that is physically accessible, a school system that provides appropriate academic modifications, or a town that has health services, many children with disabilities will no longer be able to successfully function within their own communities. Disability is thus both *magnified* and *amplified* through the interaction between the psychological or physical condition of the child with disabilities and the ongoing disruption of their environmental setting. When the postdisaster context no longer provides affordances for their disability, the individual functioning of a child with disabilities becomes more salient.

For these reasons, we see community recovery as a primary contributor to the resilience of children with disabilities following disaster. The more quickly essential services are restored within a community, the more likely children with disabilities and their families will be able to access needed supports and accommodations. Thus, the recovery of an individual child with disabilities may be closely connected to the recovery of the community macrosystem. In their community resilience framework, Norris, Stevens, Pfefferbaum, Wyche, and Pfefferbaum (2008) suggest that in order to increase their resilience to disaster, communities need to attend to their areas of greatest social vulnerability. Children with disabilities are one of these socially vulnerable populations, but we do not yet know what parts of the community infrastructure are most critical to support their resilience in the postdisaster context.

Given the dearth of empirical studies that actually explore how disaster affects children with disabilities in disaster, our review is necessarily speculative. Theoretically informed and methodologically rigorous quantitative and qualitative research is still needed to delineate the special risks and vulnerabilities that children with disabilities may experience both before and after disaster and to identify vital protective factors. Consequently, several key questions have not yet been addressed: How do various child characteristics (i.e., age, developmental level, intelligence, personality, gender, ethnicity, social class, family structure, geographic location) intersect with disability to influence the experiences of children with disabilities in disaster? How does disability type (i.e., auditory, visual, physical, cognitive) impact the ability of children with disabilities and their families to prepare for, respond to, and recover from different types of disaster? What roles do social stigma and institutional exclusion play in limiting the access of children with disabilities to emergency preparedness planning, sheltering options, therapeutic interventions, and other disaster-related resources? How are individuals within the social network of a child with a disability (e.g., parents, teachers, friends) affected by disaster and how does this influence the psychological response of the child? Are school-aged children with disabilities more likely to endure academic disruption, and if so, what are the developmental consequences of that disruption? What factors best support the resilience of children with disabilities postdisaster?

In order to design effective interventions for children with disabilities in disaster, future research needs to recognize the multilayered effects of how families, schools, communities, and societies shape the environment around children with disabilities as well as the types of adversity that they face during and following disaster (see Masten & Obradovic, 2008). In particular, we need to understand not only what makes these children vulnerable to disaster but also what can make them more resilient to disaster. Although disasters, at first glance, may appear to be what Lynch and Cicchetti (1998) refer to as *transient* risk factors, our discussion reveals them as potentially *enduring* vulnerability factors that represent ongoing conditions of inequity that place these children and their families at risk.

Given that children with disabilities often already face developmental challenges, disaster is one more risk factor that can lead to unfolding adverse consequences as these children develop.

All children, including those with disabilities, have the right to be safe and protected from natural and technological hazards and willful human-caused disasters. Certainly, children with disabilities should have equal access to disaster-related services as do their nondisabled peers, and in a just world, disability should not make them differentially at risk for injury, death, or psychological impairment following disaster.

References

Abramson, D., & Garfield, R. (2006). *On the edge: Children and families displaced by Hurricane Katrina and Rita face a looming medical and mental health crisis*. New York: Columbia University Mailman School of Public Health.

Abramson, D., Redlener, I., Stehling-Ariza, T., & Fuller, E. (2007). *The legacy of Katrina's children: Estimating the numbers of at-risk children in the Gulf Coast states of Louisiana and Mississippi*. New York: Columbia University Mailman School of Public Health.

Aldrich, N., & Benson, W. F. (2008). Disaster preparedness and the chronic disease needs of vulnerable older adults. *Preventing Chronic Disease*, *5*(1), 1–6.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th Ed., Text Rev.). Washington, DC: Author.

Americans with Disabilities Act. (1990). 42 U.S.C.A. Section 12101 et seq. (West 1993).

Appleyard, K., & Osofsky, J. D. (2003). Parenting after trauma: Supporting parents and caregivers in the treatment of children impacted by violence. *Infant Mental Health Journal*, *24*, 111–125.

Babugura, A. A. (2008). Vulnerability of children and youth in drought disasters: A case study of Botswana. *Children, Youth and Environments*, *18*(1), 126–157.

Barile, M., Fichten, C., Ferraro, V., & Judd, D. (2006). Ice storm experiences of persons with disabilities: Knowledge is safety. *The Review of Disability Studies*, 2(3), 35–48.

Bartlett, S. (2008). The implications of climate change for children in lower-income countries. *Children, Youth and Environments*, *18*(1), 71–98.

Breton, J., Valla, J., & Lambert, J. (1993). Industrial disaster and mental health of children and their parents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 438–445.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*, 568–586.

Bystritsky, M. D., Vapnik, R., Maidment, K., Pynoos, R. S., & Steinberg, A. M. (2000). Acute responses of anxiety disorder patients after a natural disaster. *Depression and Anxiety*, *11*, 43–44.

Casserly, M. (2006). Double jeopardy: Public education in New Orleans before and after the storm. In C. Hartman & G. D. Squires (Eds.), *There is no such thing as a natural disaster: Race, class, and Hurricane Katrina* (pp. 197–214). New York: Routledge.

Christ, G. H., & Christ, T. W. (2006). Academic and behavioral reactions of children with disabilities to the loss of a firefighter father: The New York World Trade Center attack 9/11/01. *The Review of Disability Studies*, 2(3), 68–77.

Cicchetti, D., & Lynch, M. (1993). Toward an ecological/transactional model of community violence and child maltreatment: Consequences for children's development. *Psychiatry*, *56*, 96–118.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, *84*, 242–261.

Dash, N., & Gladwin, H. (2007). Evacuation decision making and behavioral responses: Individual and household. *Natural Hazards Review*, *8*, 69–77.

Dubow, E. F., & Tisak, J. (1989). The relation between stressful life events and adjustment in elementary school children: The role of social support and social problem-solving skills. *Child Development*, *60*, 1412–1423.

Fothergill, A., & Peek, L. (2004). Poverty and disasters in the United States: A review of the sociological literature. *Natural Hazards*, 32, 89–110.

Fothergill, A., & Peek, L. (2006). Surviving catastrophe: A study of children in Hurricane Katrina. In Natural Hazards Center (Ed.), *Learning from catastrophe: Quick response research in the wake of Hurricane Katrina* (pp. 97–130). Boulder: Institute of Behavioral Science, University of Colorado.

Fox, M. H., White, G. W., Rooney, C., & Rowland, J. L. (2007). Disaster preparedness and response for persons with mobility impairments. *Journal of Disability Policy Studies*, *17*(4), 196–205.

Garfinkel, I., Kaushal, N., Teitler, J., & Garcia, S. (2005). Vulnerability and resilience: New Yorkers respond to 9/11. In N. Foner (Ed.), *Wounded city: The social impact of 9/11* (pp. 28–75). New York: Russell Sage Foundation.

Green, B. L., Korol, M., Grace, M., Vary, M. G., Leonard, A. C., Goldine, C. G., et al. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the Academy of Child and Adolescent Psychiatry*, *30*, 945–951.

Hemingway, L., & Priestly, M. (2006). Natural hazards, human vulnerability, and disabling societies: A disaster for disabled people? *The Review of Disability Studies*, 2(3), 57–67.

Hewitt, K. (1997). *Regions of risk: A geographical introduction to disasters*. Boston: Addison Wesley Longman.

Hewitt, K. (2007). Preventable disasters: Addressing social vulnerability, institutional risk, and civil ethics. *Geographischs Rundschau International Edition, 3*(1), 43–52.

Institute of Medicine. (2001). *Neurological, psychiatric, and developmental disorders: Meeting the challenge in the developing world*. Washington, DC: National Academy Press.

International Federation of Red Cross and Red Crescent Societies. (2007). *World disasters report 2007: Focus on discrimination*. Bloomfield, CT: Kumarian Press.

Kailes, J. I., & Enders, A. (2007). Moving beyond 'special needs': A function-based framework for emergency management and planning. *Journal of Disability Policy Studies, 17*(4), 230–237.

Kinne, S., Patrick, D. L., & Doyle, D. L. (2004). Prevalence of secondary conditions among people with disabilities. *American Journal of Public Health, 94*(3), 443–445.

La Greca, A. M. (2001). Children experiencing disasters: Prevention and intervention. In J. Hughes, A. La Greca, & J. Conoley (Eds.), *Handbook of psychological services for children and adolescents* (pp 195–224). New York: Oxford University Press.

La Greca, A. M., Silverman, W. K., Vernberg, E. M., & Roberts, M. C. (Eds.). (2002). *Helping children cope with disasters and terrorism*. Washington, DC: American Psychiatric Association.

Lynch, M., & Cicchetti, D. (1998). An ecological-transactional analysis of children and contexts: The longitudinal interplay among child maltreatment, community violence, and children's symptomatology. *Development and Psychopathology, 10*, 235–257.

Masten, A. S. (2001). Ordinary magic. Resilience processes in development. *The American Psychologist, 56*, 227–238.

Masten, A. S., & Obradovic, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society, 13*(1), 9.

McMillen, C., North, C., Mosley, M., & Smith, E. (2002). Untangling the psychiatric comorbidity of posttraumatic stress disorder in a sample of flood survivors. *Comprehensive Psychiatry, 43*(6), 478–485.

Mitchell, T., Haynes, K., Hall, N., Choong, W., & Oven, K. (2008). The role of children and youth in communicating disaster risk. *Children, Youth and Environments, 18*(1), 254–279.

Morrow, B. H. (1999). Identifying and mapping community vulnerability. *Disasters, 23*(1), 1–18.

National Organization on Disability. (2005). *Report on special needs assessment for Katrina evacuees (SNAKE) project*. Washington, DC: National Organization on Disability.

Neumayer, E., & Plümper, T. (2007). The gendered nature of natural disasters: The impact of catastrophic events on the gender gap in life expectancy, 1981–2002. *Annals of the Association of American Geographers, 97*(3), 551–566.

Norris, F. H., Friedman, M. J., Watson, P. J., Byrne, C. M., Diaz, E., & Kaniasty, K. (2002). 60,000 Disaster victims speak: Part I. An empirical review of the empirical literature, 1981-2001. *Psychiatry, 65*(3), 207–239.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche Karen, F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology, 41*, 127–150.

O'Keefe, P., Westgate, K., & Wisner, B. (1976). Taking the naturalness out of natural disasters. *Nature, 260*, 566–567.

Osofsky, J. D. (1995). The effects of exposure to violence on young children. *American Psychologist, 50*, 782–788.

Osofsky, J. D., Osofsky, H. J., & Harris, W. W. (2007). Katrina's children: Social policy considerations for children in disasters. *Social Policy Report, 21*(1), 3–18.

Peek, L. (2008). Children and disasters: Understanding vulnerability, developing capacities, and promoting resilience. *Children, Youth and Environments, 18*(1), 1–29.

Peek, L., & Fothergill, A. (2008). Displacement, gender, and the challenges of parenting after Hurricane Katrina. *National Women's Studies Association Journal, 20*(3), 69–105.

Penrose, A., & Takaki, M. (2006). Children's rights in emergencies and disasters. *The Lancet, 367*, 698–699.

Phillips, B. D., & Morrow, B. H. (2007). Social science research needs: Focus on vulnerable populations, forecasting, and warnings. *Natural Hazards Review, 8*(3), 61–68.

Picou, J. S., & Marshall, B. K. (2007). Social impacts of Hurricane Katrina on displaced K-12 students and educational institutions in coastal Alabama counties: Some preliminary observations. *Sociological Spectrum, 27*, 767–780.

Rooney, C., & White, G. W. (2007). Narrative analysis of a disaster preparedness and emergency response survey from persons with mobility impairments. *Journal of Disability Policy Studies, 17*(4), 206–215.

Rowland, J. L., White, G. W., Fox, M. H., & Rooney, C. (2007). Emergency response training practices for people with disabilities. *Journal of Disability Policy Studies, 17*(4), 216–222.

Scotti, J. R., Stevens, S., Cavender, A., Morford, M., Jacoby, V., Freed, R., et al. (2007). *Response of persons with mental retardation/developmental disabilities to emergency situation: Implications for disaster preparedness*. Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, Baltimore.

Smart, J. (2001). *Disability, society, and the individual*. Gaithersburg, MD: Aspen.

Spence, P. R., Lachlan, K., Burke, J. M., & Seeger, M. W. (2007). Media use and information needs of the disabled during a natural disaster. *Journal of Health Care for the Poor and Underserved, 18*, 394–404.

Stough, L. M. (2009). The effects of disaster on the mental health of individuals with disabilities. In Y. Neria, S. Galea, & F. H. Norris (Eds.), *Mental health consequences of disasters* (pp. 264–276). New York: Cambridge University Press.

Stough, L. M., & Sharp, A. N. (2007). *The recovery of individuals with disabilities following Hurricane Katrina*. Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, Baltimore.

Tierney, K. J., Petak, W. J., & Hahn, H. (1988). *Disabled persons and earthquake hazards*. Boulder, CO: Institute of Behavioral Science, University of Colorado.

United Nations Human Settlements Programme. (2007). *Enhancing urban safety and security: Global report on human settlements*. London: Earthscan.

United Nations International Children's Emergency Fund. (2007). *Promoting the rights of children with disabilities*. Florence, Italy: UNICEF Innocenti Research Center.

U.S. Census Bureau. (2005). *Survey of income and program participation, June–September 2005*. Retrieved May 8, 2010, from http://www.census.gov/hhes/www/disability/sipp/disab05/d05tb4.pdf

U.S. Census Bureau. (2006). *2006 American community survey data*. Washington, DC: U.S. Census Bureau. Retrieved February 1, 2008, from http://www.census.gov/acs/www/

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2009). *28th annual report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2006, Vol. 1*. Washington, DC: Author.

U.S. Federal Emergency Management Agency and Department of Homeland Security Office for Civil Rights and Civil Liberties. (2008). *Interim emergency management planning guide for special needs populations*, Version 1.0. Washington, DC: U.S. Federal Emergency Management Agency and Department of Homeland Security Office for Civil Rights and Civil Liberties.

Van Willigen, M., Edwards, T., Edwards, B., & Hessee, S. (2002). Riding out the storm: Experiences of the physically disabled during Hurricanes Bonnie, Dennis, and Floyd. *Natural Hazards Review*, *3*(3), 98–106.

Weissbecker, I., Sephton, S. E., Martin, M. B., & Simpson, D. M. (2008). Psychological and physiological correlates of stress in children exposed to disaster: Review of current research and recommendations for intervention. *Children, Youth and Environments*, *18*(1), 30–70.

White, B. (2006). Disaster relief for deaf persons: Lessons from Hurricanes Katrina and Rita. *The Review of Disability Studies*, *2*(3), 49–56.

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2004). *At risk: Natural hazards, people's vulnerability, and disasters*, 2nd ed. New York: Routledge.

World Health Organization. (2001). *International classification of functioning, disability, and health*. Geneva, Switzerland: Author.

World Health Organization. (2005). *Disability, including prevention, management, and rehabilitation*. Geneva, Switzerland: Author.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1271–1286

Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study

Alastair Ager, Lindsay Stark, Bree Akesson, and Neil Boothby
Columbia University

Thirty specialists in humanitarian work supporting the care and protection of children in crisis settings completed a 3-phase Delphi consultation. Proposals of best practice were elicited, reviewed, and rated by participants. A high level of consensus support was reached for 55 statements. These statements emphasized utilization of existing resources, participation, and inclusivity. The influences of resilience theory, social ecology, and cultural sensitivity were clearly evident. The utilization of developmental theory could be strengthened in relation to more differentiated understanding of the operation of protective influences and conceptualization of such influences in terms of "adaptive systems." Wider research engagement by development scientists in diverse cultural settings and clear formulation of findings for practitioners and policy makers would further support evidence-based humanitarian practice.

Globally, natural and man-made disasters represent a significant risk to children's developmental progress and well-being (Boothby, Strang, & Wessells, 2006; Bryce & Boschi-Pinto, 2005; Carballo, Heal, & Horbaty, 2006; Cohen et al., 2005). Estimates place the number of children involved in armed conflict at over 300,000 (Coalition to Stop the Use of Child Soldiers, 2004). During the Mozambican civil war, 64% of children aged between 5 and 15 years experienced abduction from their families and 77% witnessed murder (Boothby, Strang, et al., 2006). An estimated half million children, one fifth of the child population, were separated from their parents as a result of the Rwandan genocide (Cantwell, 1997). The 2004 tsunami orphaned well over 20,000 children, with estimates that approaching 40% of these were under the age of 5 (Carballo et al., 2006). Over 5,000 children were left without an adult caregiver in the immediate aftermath of the Sichuan earthquake (ReliefWeb, 2008). It is estimated that over 12 million children have been orphaned by AIDS in Sub-Saharan Africa (UNAIDS, 2004).

These global crises pose significant risks to children's well-being and developmental progress. Not only do children face increased vulnerability to their immediate physical well-being, but childhood development is often also compromised with serious risks for long-term growth and survival. These crises often result in decreased health and nutrition status among children, seriously impacting the important childhood years of physical and mental development. Childhood malnutrition has consistently been linked to growth stunting, cognitive deficits, and decreased school attendance (Berkman, Lescano, Gilman, Lopez, & Black, 2002; Hoddinott & Kinsey, 2001; Martorell, Rivera, Kaplowitz, & Pollitt, 1992; Mendez & Adair, 1999; Walker, Chang, Powell, & Grantham-McGregor, 2005) as well as increased morbidity and mortality rates (Bryce & Boschi-Pinto, 2005; Toole & Waldmen, 1997). The lack of primary caretakers' capacity to meet children's needs for care and support—be they missing, wounded, ill, or otherwise unavailable to provide their usual level of supervision—represents another consistent threat to developmental well-being (Levy-Shiff, Hoffman, &

The authors acknowledge the contributions of Deborah Tsuchida, Andrew McMahan, Ann Warner, Elisabeth Berger, and Mark Orr in the implementation of work described, and commentary on drafts of the manuscript provided by Ezer Kang, Pamela King, and Wendy Ager. The work was completed as part of the Care and Protection of Children in Crisis-Affected Countries (CPC). Initiative funded by USAID's Displaced Children and Orphans Fund, the Oak Foundation, and the United States Institute for Peace. Statements made in this article are the views of the authors alone, and do not constitute the policy of any of these agencies.

The work was conducted with respect to the Columbia University Medical Center Institutional Review Board determination IRB-AAAB7134.

Correspondence concerning this article should be addressed to Alastair Ager, Program on Forced Migration and Health, Mailman School of Public Health, Columbia University, 60 Haven Avenue, B-4, New York, NY 10027. Electronic mail may be sent to aa2468@columbia.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0017

Rosenthal, 1993; Williams, Hyder, & Nicoli, 2005). The importance of such factors as cognitive stimulation, caregiver sensitivity, responsiveness, and caregiver affect in supporting children's cognitive and socioemotional development highlights the risks to children in contexts of disaster of disruption to the protective environment of supportive caretakers (Brazelton & Sparrow, 2006; Walker et al., 2007; World Health Organization, 2004). Separation from caretakers also leaves children susceptible to exploitation and abuse (Hepburn, 2006).

From an ecosystems perspective, disaster and conflict alike erode the social fabric of communities, undermining the capacities and institutions that provide developmental support to children across varying ''ages and stages'' beyond immediate caregivers (Ahearn, Loughry, & Ager, 1999; Boothby, Crawford, & Halperin, 2006; Carballo et al., 2006). Crises disrupt the day-to-day activities that fill a child's life and which allow children to comfortably and safely explore and express themselves. In crisis settings, formal and nonformal systems of learning may be destroyed, seriously interrupting the opportunity for children's cognitive stimulation and critical thinking. Abduction results in disruption of schooling, social marginalization, and delayed advancement of economic livelihoods (Annan, Blattman, & Horton, 2006). In conflict settings in particular, disruption of civic and religious institutions may significantly threaten children's moral development and evolving spiritual practices (Wessells & Strang, 2006). All this represents major disruption of key adaptive systems (Masten & Obradovic, 2008) supporting developmental outcomes.

Given these threats, response to the well-being, protection, and developmental needs of children is a major component of contemporary humanitarian intervention in crisis situations. In 2007, UNICEF, the United Nations' lead children's agency, launched more than 50 distinct funding appeals (totaling US$874 million) to support humanitarian interventions addressing the protection needs of vulnerable children and women (UNICEF, 2008). To guide such work, a Child Protection Working Group (CPWG) has recently been established within the Protection Cluster Working Group, the principal coordination mechanism for United Nations and international nongovernmental organization (NGO) work in humanitarian crises (OCHA, 2007).

However, the evidence base for effective response in such circumstances is weak. The lack of rigorous empirical studies documenting the effectiveness of specific programmatic interventions is a consistent theme in the literature across a wide range of contexts (Ahearn, 2000; Boothby, Strang, et al., 2006; Brooks, 2005; Wessells, 2006; Williamson & Cripe, 2002). Repeated calls to implement more empirical evaluations of child protection and related programming have been made (Ager, Boothby, & Wessells, 2007; Leaning, 2001; Hofmann, Roberts, Shoham, & Harvey, 2004; Weinstein, 2005), but the complexity of crisis settings and the urgency of humanitarian response conspire against robust study design. Given limited human and financial resources, the rapid roll-out of services and interventions often takes precedence over research and evaluation efforts.

There are a number of recent examples of field data collection in crisis situations that suggest that these challenges are beginning to be successfully addressed (e.g., Annan et al., 2006; Betancourt, Pochan, & de la Soudiere, 2005). However, there are alternative strategies to establish a core knowledge base to guide intervention in a specific field that stand to complement such efforts. These are broadly known as *consensus methodologies* (Duncan, 2006).

Consensus methodologies build on the awareness that valuable knowledge is gained by supporting processes of reflection by professionals (Kolb, 1984; Schon, 1983). Experiential, reflective learning is an important source of knowledge for developing professional expertise. Humanitarian professionals have made much use of such reflection (Wood, Apthorpe, & Borton, 2001), developing a strong culture of interagency consultation to develop consensus guidelines on a range of issues (e.g., IASC, 2007; The Sphere Project, 2004). However, such consultations seldom adopt the rigor of established consensus methodologies, such as Delphi, which provide a highly structured means to distill key lessons learned by experienced practitioners across a range of settings (Bowling, 2002; Duncan, 2006; Jones & Hunter, 1995; Murphy et al., 1998; Walker & Selfe, 1996). Crucially, such methods provide a means to control the social and political processes that can otherwise distort the distillation of knowledge involved in developing consensus (essentially the under- or overweighting of specific elements of evidence as a result of status or other forms of influence).

The current study was an attempt to identify, through a Delphi consensus methodology, the core knowledge base that senior practitioners in the field of care and protection of children in crisis settings considered to constitute ''best practice'' at this time. We considered that seeking to determine such a knowledge base would serve two valuable functions. First, explicitly identifying areas where there is consensus and where it is currently lacking provides a sound basis for programmatic and policy reflection within the field. This can usefully involve both dissemination of program guidance (where there is established agreement) and focused debate (in areas of disagreement). Second, defining the knowledge base on which the field currently operates allows consideration of the extent to which this foundation effectively reflects relevant insights of developmental science and, further, the potential contribution of developmental research to enhancing such knowledge.

Method

Participants

We determined that the expertise most influential on current international humanitarian response to the care and protection needs of children was held by practitioners working with major agencies active in the field, with experience across a diverse array of settings, and accountable at a senior level (allowing significant influence on practice). Potential participants were therefore defined with respect to four criteria: (a) holding the most senior position in child protection within a leading donor agency working in the children in crisis field, (b) holding the most senior position within an international network or forum regarding children in crisis situations, (c) holding the position of senior child protection adviser or equivalent or above within specialized children's services within one of a specified number of "leading agencies" (intergovernmental organizations and international NGOs) working with children in crisis, and (d) having been employed/contracted by one or more of these "leading agencies" for technical/advisory work, leading to the production of five or more technical reports. "Leading agencies" were defined with respect to positions advertised and projects listed on the ReliefWeb and IRIN Web sites within the preceding year (a criterion that served to include all governmental and nongovernmental members of the CPWG of the Protection Cluster Working Group, the body mandated by the UN to coordinate work related to the care and protection of children in the context of disaster and associated humanitarian crisis).

Seventy-seven potential participants were identified through Internet searches and telephone enquiry as fulfilling one or more of the above criteria. Contact by e-mail or telephone was established for 52 of these. Thirty-eight consented to enroll within the study and 31 completed Phase 1. Thirty participants completed Phases 2 and 3 of the study. Completing participants (21 women and 9 men) were drawn from all eligibility criteria, although there was a trend approaching significance ($\chi^2 = 6.90$, $df = 3$, $p = .075$) for more participants enrolled under latter criteria to complete the study (completion rates 10%, 17%, 30%, and 71%, respectively, for Criteria 1–4). Participants were thus more representative of technical and program design expertise than those responsible for decision making on funding.

Procedure

The study adopted a three-phase Delphi design (Duncan, 2006; Green, Jones, Hughes, & Williams, 1999; Hasson, Keeney, & McKenna, 2000). Additional phases may promote higher levels of consensus, but where attrition of participants is considered a risk—as it was here with an Internet-based consultation with senior practitioners—this may be at the cost of representativeness (Duncan, 2006; Hasson et al., 2000; Powell, 2002; Schmidt, 1997). The study was conducted between May 2006 and April 2007. Participants were regularly contacted by study administrators during this period to inform them on progress of the study and, where necessary, encourage timely completion of responses to allow the study to continue to the next phase.

Phase I: Elicitation of proposed best practice. Those consenting to enroll in the study were invited to propose, via e-mail, "best practice" in the care and protection of children in crisis-affected settings. Participants proposed such practice by submitting statements in a specified format. By the end of this initial consultation period, a total of 255 statements summarizing best practice had been submitted. A manageable, consolidated listing of statements for review in the next phase of the study was developed using successive thematic analysis (Silverman, 2001; see Figure 1 for a schematic representation of the stages followed).

First, three researchers independently reviewed a subsample of statements to develop preliminary categorical codes, which were then consolidated into the broad categories listed in Table 1. All proposed statements were then grouped into these categories. Second, two researchers independently analyzed statements within each category, grouping them to define constituent themes. An agreed set of themes within each category was then negotiated. Third, review statements were compiled to represent each of these themes. These review

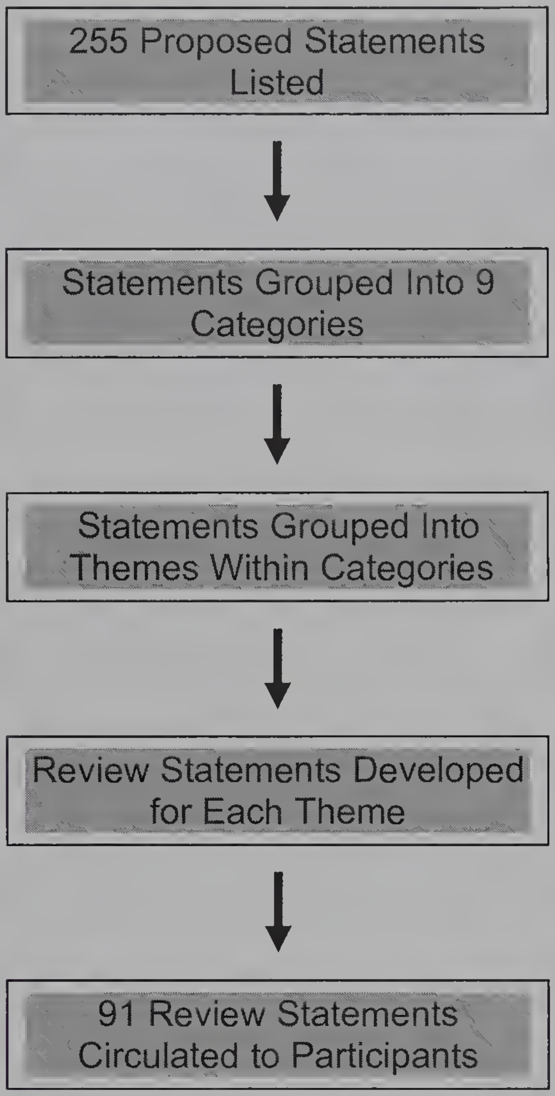

255 Proposed Statements Listed
Statements Grouped Into 9 Categories
Statements Grouped Into Themes Within Categories
Review Statements Developed for Each Theme
91 Review Statements Circulated to Participants

Figure 1. Stages of thematic analysis to identify review statements.

statements were derived from listed statements, focusing on recurrent elements within each theme. Again, two researchers independently compiled statements and then negotiated agreed wording (targeting accessibility and lack of ambiguity). At least one statement was compiled for each theme; additional statements were compiled where themes represented five or more proposed statements and the initial statement drafted did not adequately represent key elements reflected in the theme. This resulted in a listing of 91 review statements.

Phase II: Preliminary ratings of proposed best practice. In Phase II, this listing was presented to all participants with the request to rate their level of agreement with each statement as *strongly agree, agree, undecided, disagree,* or *strongly disagree.* For each statement, there was also space for participants to enter comments related to their ratings.

Phase III: Finalized ratings of proposed best practice. In Phase III, the composite ratings of statements from Phase II were fed back to individuals, providing them with the opportunity to amend—or retain—their own ratings given knowledge of others' judgments. Participants were presented with a listing of statements indicating the percentage of participants who had marked *strongly agree, agree, undecided, disagree,* and *strongly disagree* for each. They were also shown the rating that they themselves had given for each statement and the comments provided on that statement by all participants. For each statement they then marked their final rating and made further comments on the statement as desired.

Results

Figure 2 summarizes the distribution of ratings across the 91 statements that emerged at the end of the third phase of consultation. As was to be expected with statements proposed as representing perceived "best practice," the distribution of ratings was positively skewed, with a median rating in the *strongly agree* range. Mean rating was 1.63, more than 3 times the standard error (0.395) from the center point of the 5-point rating scale.

Level of consensus was defined as the percentage of participants indicating *agreement* or *strong agreement* with a statement. Figure 3 displays the range of consensus support achieved across statements. Following emerging convention in Delphi studies (Beattie & Mackway-Jones, 2004; Brown, Crawford, Carley, & Mackway-Jones, 2006; Crawford, Mackway-Jones, Russel, & Carley, 2004; Duncan, 2006), 90% and 80% consensus were considered as potential thresholds to judge a statement as consensually agreed upon. With the gradient of this curve steepening beyond the 90% threshold, and then dropping sharply after the 80% threshold, the 90% threshold was taken as defining "clear consensus", with statements in the 80%–89% range designated as "approaching consensus."

Table 1 summarizes the level of consensus among participants for each statement. Statements are listed in descending order of consensus within each thematic grouping, and referenced by statement number and statement summary (listing of full statements is available at http://www.cpclearningnetwork.org).

Consensus is the key indicator for the interpretation of findings, but two other indices were calculated to support interpretation. Intensity ratios are commonly used in policy research to quantify the strength of opinion (including minority opinion) on Likert-type scales (Inglehart, 1995; Kemp & Burt, 2002; O'Hara & Stagl, 2002). Intensity ratios were

Table 1

Consensus Ratings of Reviewed Best Practice Statements

		Consensus (%)	AI	DI
Agency Strategy				
14	We must establish a "culture of learning" among agencies	100	0.60	0.00
11	Good practice requires attention to "Do No Harm" imperative	97	0.79	1.00
20	Child protection experts should be rapidly deployed to work with specialists in other sectors	97	0.78	0.00
13	Agency codes of conduct must be signed and enforced	96	0.86	0.00
17	Staff security must be considered at all times	96	0.82	0.00
18	Those delivering services to children must receive on-going support and training	96	0.57	0.00
4	Care and protection of children needs to be seen as concern of all agencies, working in an integrated, and coordinated manner	93	0.68	1.00
7	Coordination between agencies brings great benefits	93	0.54	0.00
9	We need to work at many levels, not just providing direct services	92	0.77	0.50
22	Advocacy needs to be data driven and based on program learning	89	0.64	0.00
3	Capacity building of local staff and partner organizations should be a component of all interventions	87	0.77	0.23
12	Consensus on definitions that relate to child care and protection is needed	87	0.39	0.77
8	It is good practice to build government capacity and avoid creating parallel systems	86	0.48	0.27
10	Documents, translated into all relevant languages, should be broadly distributed	83	0.28	0.59
15	Agencies need to balance investments between resident, IDP, and refugee populations and on both sides of civil wars	82	0.44	0.81
5	A rights-based approach must be continuously promoted	74	0.36	0.50
21	Failure to educate donors contributes to poor practice and lack of long-term commitment	71	0.35	0.38
6	Necessary protection and care actions are often broadly similar across situations. Good practice is having one set of guiding principles that are agreed to by all those involved in the care and protection of children	70	0.14	0.33
16	Effective protection programming must include nonviolent conflict resolution	61	0.30	0.36
19	Interventions need to be scalable	61	0.52	0.72
2	Emergency needs should be addressed independently of long-term strategy	33	0.39	0.85
Community Engagement and Participation				
27	In enabling community mobilization, it is vital to identify and include different community subgroups	100	0.77	0.00
23	It is essential to not ignore or undermine existing or traditional mechanisms	94	0.71	0.50
25	It is best practice to take a participatory approach with children and youth guiding the design, implementation, and evaluation of programs	93	0.43	0.50
28	It is good practice to conduct a participatory assessment with affected people in coordination with other assessment processes	93	0.68	0.00
29	We should support and foster development of preexisting child protection support networks	90	0.63	0.00
26	Children themselves are a neglected resource in child protection, and should be seen as social actors not as passive recipients of care and protection services	87	0.54	0.23
24	Best practice involves community-designed, community-managed, and community-owned projects	77	0.13	0.26
Understanding Children's Needs				
31	Programming should be inclusive and reach out to a range of affected children	100	0.47	0.00
35	Long-term strategies are required for youth who have missed education and who need to become economically active	100	0.50	0.00

Table 1
Continued

		Consensus (%)	AI	DI
37	First line action in protection is to limit exposure of children to traumatic events and provide activities that create sense of normality	96	0.45	0.00
32	Disaggregated data (by age, sex, etc.) are needed to ensure effective and appropriate programming	94	0.68	0.00
39	Agencies need to develop their skills for working at the community level in supporting families in caring for children	94	0.50	0.00
38	Children with extreme reactions to loss, displacement, etc. and those with severe mental illness should also be considered in program design and service delivery	93	0.46	0.00
36	Planning and delivery of services needs to be child-focused	90	0.33	0.00
30	We need to avoid categorizing and labeling children to avoid stigmatization and jealousy	76	0.68	0.13
34	Program assessments and interventions should consider children holistically and not in terms of sectors	73	0.38	0.11
33	Using the term "vulnerable" should be avoided	27	0.63	0.38
Monitoring, Evaluation, and Research				
40	We need to develop an evidence base of what constitutes effective child care and protection	100	0.83	0.00
46	We need strict ethical protocols for collecting information from children	100	0.72	0.00
43	It is good practice to evaluate program impact using a mixture of qualitative and quantitative data	97	0.72	0.00
45	Evaluation should consider project impact on social dynamics and power structures within communities	93	0.75	0.00
41	Effective programs must be able to demonstrate results with respect to clearly specified indicators and outcomes	93	0.48	0.50
42	Monitoring data should be used to adjust project implementation and evaluation data should be used to inform future planning	90	0.58	0.00
44	To assess impact we need rigorous data	69	0.20	0.32
Gender				
51	We need to provide "girl friendly" reproductive health and GBV services	100	0.82	0.00
48	Programs need to proactively reach out to engage girls	96	0.74	0.00
50	It is important to use local cultural resources as one means of supporting war-affected girls; it is equally important to be critical about practices which may violate human rights	96	0.61	1.00
52	There is a need for programming that targets and educates men and boys about gender roles and sexual responsibility	96	0.78	0.00
47	A gender perspective needs to be applied to all programming	93	0.77	0.50
53	We should avoid segregating girls into "appropriate for gender" skills training	86	0.42	0.21
49	We should support "positive discrimination" as one criterion to ensuring equitable treatment of girls	57	0.32	0.32
Separated Children				
58	We need to put strategies in place to prevent the separation of children	100	0.68	0.00
55	Understanding cultural norms can help protect children from exploitation	96	0.67	1.00
57	Thorough documentation is necessary for separated children before and during a crisis	96	0.41	0.00
62	It is best practice to monitor protection and well-being of separated children living with extended family or in foster care	96	0.78	0.00
59	Mobilization and mediation efforts are needed to facilitate the process of reintegrating separated children	93	0.61	0.00
60	Institutional care must be avoided whenever possible	89	0.61	0.00
56	Organizations need to be active in tracing separated children	85	0.46	0.50

Table 1
Continued

		Consensus (%)	AI	DI
61	Adoption should be avoided in crisis situations unless the child will not be cared for by extended family and may be adopted within their own community	79	0.37	0.00
54	It is best practice to offer health screening, health care, and shelter to separated children	78	0.24	0.32
Children Associated with Fighting Forces				
64	Effective reintegration programs support former children associated with fighting forces and also other vulnerable children	100	0.61	0.00
69	There is a need for planned reintegration from a long-term perspective with recognition of ongoing needs	100	0.59	0.00
71	Efforts to support children formerly associated with fighting forces are most effective and sustainable when based on their strengths and resources	100	0.46	0.00
75	Child protection must be addressed and prioritized within military and peacekeeping operations	100	0.82	0.00
76	Agencies need to understand how cultural, social, and political processes affect recruitment and use of children in armed forces	96	0.67	0.00
66	Separate DDR processes are needed for children and adults	93	0.61	0.50
68	Effective reintegration programs include supports relating to family mediation, health, education, livelihood, nonviolent conflict resolution, spiritual well-being, and community protection	93	0.61	0.00
70	Family tracing and reunification should start as soon as children are released from armed forces	93	0.44	0.50
72	Children should be supported in leaving fighting forces at any time during conflict	93	0.85	0.00
74	Substance abuse prevention, education, assessment, and treatment should be part of reintegration programs	93	0.23	0.00
77	Protection must be addressed in transitional justice processes	89	0.72	0.00
67	Girls and girl-mothers need to be included in formal and informal DDR processes	86	0.79	0.00
78	Cultural mechanisms for reintegration are highly effective and should generally be utilized	85	0.59	0.20
65	It is best practice to not provide cash settlements to children associated with fighting forces	78	0.50	0.18
63	Children associated with fighting forces should be provided with targeted support during reintegration	75	0.24	0.72
73	Vocational training can alienate rather than reintegrate a target group of children by drawing them away from other community activities	27	0.15	0.21
Schooling and Education				
81	Specific strategies need to be put in place to engage girls in education and training activities	100	0.50	0.00
79	The re-establishment of schooling is a key protective measure for children	97	0.54	1.00
84	Youth interventions need to go beyond formal schooling to include nonformal activities	92	0.64	0.00
83	Provision of safe spaces, recreational opportunities, and youth clubs can be valuable activities	90	0.32	0.00
80	Payment and training of school teachers is required to establish quality education for children	89	0.36	0.36
85	Parental education is critical to program success	64	0.22	0.39
82	Developing school management committees builds longer-term capacity within communities	61	0.48	0.08
86	Organizations should take education programs to scale only when they have logistics and trained personnel	47	0.08	0.27

Table 1
Continued

		Consensus (%)	AI	DI
Livelihoods				
90	Agencies should make income generating activities accessible for girls	97	0.67	0.00
88	Steps should be taken to support livelihoods of families in which children are thought to be particularly vulnerable	96	0.59	0.00
87	(Re)establishment of livelihoods plays a fundamental role in child protection	93	0.69	0.50
91	Livelihoods programs need to target youth	92	0.23	0.00
89	It is essential that vocational training and related skills development programs are based on sound market analysis	89	0.76	0.36
92	Skills training programs need to include job placement, start-up funding or microcredit, and follow-up to be effective	86	0.37	0.21

Note. 90%+ participants indicating agreement or strong agreement. 80%–89% participants indicating agreement or strong agreement. AI = agreement intensity; DI = disagreement intensity.

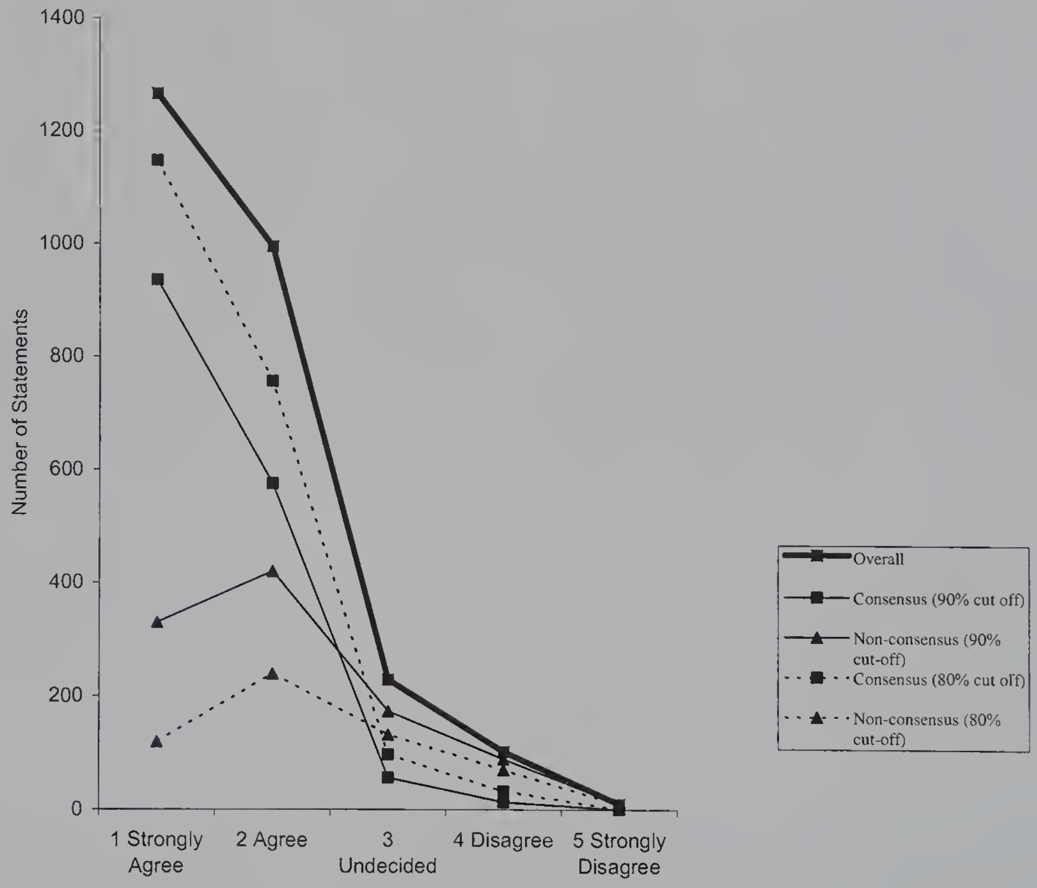


Figure 2. Frequency distribution of statement ratings (with differing "cutoff" thresholds for consensus).

therefore calculated to identify profiles of responding to particular statements that differed from the general pattern of response indicated in Figure 2. Agreement intensity (AI) was defined as the number of participants indicating strong agreement with a statement over the number of respondents indicating agreement or strong agreement with that statement. A score above 0.5 indicates that the pattern of agreements being more frequent than strong agreements (indicated by Figure 2 to be the modal pattern for nonconsensus statements) was not adhered to for this item. Disagreement intensity (DI) was defined as the number of participants indicating disagreement or strong disagreement with a statement over the number of respondents not agreeing (strongly or otherwise) with that item. It is a measure of the extent to which nonagreement involves genuine disagreement rather than uncertainty. A score of above 0.5 indicates that the modal pattern indicated by Figure 2 of responses of "undecided" being more frequent than explicit disagreement was not adhered to for this item.

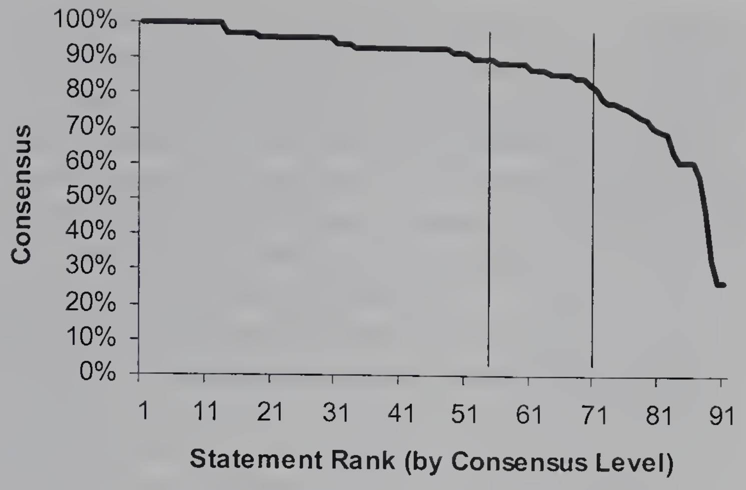


Figure 3. Levels of consensus support for statements (showing number of statements included by 90% and 80% thresholds).

Discussion

Constraints

As with all Delphi studies, the characteristics of the expert sample constrain the extent of generalization of findings. Here, the use of the Internet for identification of potential participants and of selection criteria related to "leading agencies" in the field of child protection in crisis settings provided a robust basis for inclusion of those shaping global debate in the field, but significantly underrepresents the potential contribution to such debates by "southern" child protection practitioners not visible by such means. Further, the high level of attrition of those participants meeting the inclusion criteria through their role with donor agencies urges caution in generalizing consensus to this constituency. For those meeting other inclusion criteria; however, retention of participants was high for Internet-based consensus studies (Elwyn et al. 2006; Hasson et al., 2000) and may be seen to fairly represent the technical and design expertise that shapes current programming.

The other major constraint related to the study design concerns the complex relation between statements about "best practice" and program practice in real settings. Statements were worded in a manner that sought to articulate concrete practice but, given the importance of context in determining details of implementation, they may often be seen as articulation of principle rather than defining practice per se. This leaves the statements open to differential interpretation across settings. Although the term best practice is used here to describe the intended focus of statements, the fact that such statements often articulate principles that remain open for somewhat differential implementation needs to be acknowledged.

Statements Securing Clear Consensus Support

The study identified 55 statements that secured clear consensus support (90% and above), which may be seen as representing a core knowledge base underpinning international response to child care and protection needs in crisis situations. A number of these statements reflect principles that have become established in guiding humanitarian efforts more widely including those related to the "do no harm" principle, agency coordination, staff codes of conduct and community participation (e.g., IASC, 2007; The Sphere Project, 2004). The majority, however, reflect understandings of the determinants of child well-being and development and, particularly, means of promoting this in contexts of crisis.

Participants share a view of the interconnectedness of child well-being and development with a wide range of social, cultural, and economic factors (Statements 4, 55, 76, and 87). Crises may directly expose children to trauma and loss (Statements 37 and 38), but the erosion of community resources also critically undermines well-being (Statements 23, 58, and 76). Targeted programming is supported in such areas as reproductive health services for girls (51), livelihoods programs for youth (91), documentation and tracing for separated children (57), and demobilization, disarmament and reintegration provision for war-affected youth (66). But the focus is generally not on direct service provision (Statement 9). The major emphasis is rather on utilizing, and where necessary rebuilding, community capacities (Statements 23, 29, 39, 50, 55, 59, and 88), and institutions (Statements 62, 79, 81, and 83). This involves a range of potential "adaptive systems," including kinship networks, schools, religious associations, and local healing rituals. Significant attention is paid to the manner in which interventions are planned and implemented. Active participation of children and youth (Statements 25 and 36) and the wider community (Statements 27 and 28) are considered crucial, and the inclusivity of programming (Statements 31, 32, 48, and 51) is consistently emphasized.

Statements Not Securing Clear Consensus Support

The 46 statements that failed to receive strong consensus fall into three major categories. First, there were 16 statements that secured agreement from between 80% and 89% of participants, and were considered "approaching consensus." These items generally appeared to fall short of clear consensus, either because they were considered somewhat conditional upon context (e.g., 3, 8, 10, 15, and 53) or referred to technical practices (e.g., 56, 67, and 77) that were not clearly understood by some participants.

Of rather greater interest conceptually—given the processes by which minority opinion can come to shape consensus over time (Moss & Schutz, 2001)—was a second group of items for which there was a lack of consensus but with some strong support (as indicated by an AI score of 0.50 or above, indicating that those in strong agreement equaled or outnumbered those in agreement). Four statements met this criterion: Statement 19 (regarding scalability), Statement 30 (regarding stigma), Statement 33 (regarding avoidance of the term ''vulnerable''), and Statement 65 (regarding cash settlements to children associated with fighting forces). The first of these statements was unique in attracting both a high AI and a high DI score, and is considered separately in the next section. The latter statements share a core concern: the danger of stigma and community disruption when inclusion criteria for programs are based on specific needs or exposure to specific experiences. Comments from participants suggested that these statements did not receive consensus support because of concern over targeting services. A caucus of participants was highly sensitive to the potential negative impact on community dynamics of explicitly discriminating between children on the basis of externally defined categories of ''vulnerability,'' suggesting: ''*vulnerable* is about the worst word to enter the vocabulary of development'' and ''vulnerability is a product of situations, not individuals.'' Given the prominent role of vulnerability in many conceptualizations of developmental risk (Daniel, Wassell, & Gilligan, 1999; Masten & Gewirtz, 2006), the use and application of the term in humanitarian settings clearly warrants further exploration.

A third grouping of statements not only failed to reach consensus but also were marked by strong opposition, as indicated by a DI score of 0.50 or above (those actively disagreeing with the statement equaled or outnumbered those who were undecided). Of statements failing to reach the 80% threshold level for consensus, four met this criterion (2, 5, 19, and 63). Statement 63 is essentially the converse of Statements 30, 33, and 65, and disagreement with it marks the same sensitivity to targeting support to a particular group, rather than providing communitywide supports. Comments suggest that Statement 2, addressing the balance of attention to emergency needs and long-term strategy—was ambiguous to many reviewers. By contrast, there were substantive concerns on the role of a ''rights-based approach'' (Statement 5). While a child rights framework (based on the Convention of the Rights of the Child; UNHCHR, 1989) has been adopted by a number of international humanitarian agencies, comments here principally addressed the operationalization of this approach:

> Narrow definitions happen too often. It can inspire unbalanced approaches, and can undermine capacity-building.

> . . . only if done without imposing rights or taking a moralistic stance, demeaning local practices.

Finally, as noted above, Statement 19 (interventions need to be scalable) was unique in attracting both AI and DI scores of above 0.50, suggesting strong, divided opinion on this issue. Comments ranged from the strongly supportive:

> Great interventions that do not reach the vast majority cannot be considered effective practice.

to those arguing that scalability was not an appropriate criterion for effective practice:

> The fact that interventions can't be applied beyond a small group should not necessarily preclude those interventions being carried out.

Such disagreement suggests quite different understandings of the mandate of humanitarian assistance to children in situations of crisis. Seeing child care and protection in emergencies alongside such interventions as water and sanitation provision, food distribution, and emergency health care encourages a populationwide, community-based approach, with an expectation of general coverage and thus ''scalable'' provision (Brainard & LaFleur, 2007). Intensive, focused, relatively high-cost interventions draw upon radically different programmatic assumptions.

Humanitarian Consensus and Developmental Science

Few participants—either through the provision of statements or subsequent commentary on ratings—explicitly framed best practice regarding child care and protection in crisis settings in developmental terms. However, best practice was seldom cited in the narrow terms of immediate protection during crisis alone; attention was consistently paid to the longer term trajectories of children and the impact on these of crisis experiences, and the related erosion of local capacities and institutions. Explicitly acknowledged or not, practitioners in such work are seeking to shape developmental outcomes and are making assumptions about processes of development.

Three questions follow from this. First, what ideas from developmental science does this knowledge base principally draw upon? Second, are there currently unutilized insights from developmental science that would usefully inform humanitarian response to children in crisis settings? Third, how might our findings shape the work of developmental scientists eager to impact humanitarian work with children?

The current knowledge base informing humanitarian response reflects most significantly the influence of two developmental constructs: those of *resilience* and of *social ecology*. Luthar and Cicchetti (2000) identify three distinguishing features of the construct of resilience: a focus on positive outcomes and not just negative ones, an emphasis not only on deficits but also of strengths, and consideration of the impact of the interplay of vulnerability and protective factors. These features were strongly reinforced throughout the statements achieving consensus. The concern of some participants in ascribing vulnerability to individuals rather than circumstances (Statement 33) reflects broader thinking on resilience being seen as a process rather than a personal characteristic (Luthar, Cicchetti, & Becker, 2000; Masten & Wright, 2009). In terms of Buckner, Mezzacappa, and Beardslee's (2003) distinction between inner resources (e.g., self-esteem, intellectual ability) and external resources (e.g., social support, relationship with caregiver) supporting resilience, it is clear that the latter receive much more attention in framing programmatic response. Participants generally drew much more comfortably upon social constructs than psychological ones, and although this may reflect uncertainty regarding the validity of psychological constructs across cultures (IASC, 2007), it also appears a pragmatic response given the available intervention modalities and resources in most humanitarian emergency settings.

This emphasis reflects also the influence of social ecological constructs on identified best practice. Such analysis—based on conceptualizations stemming from the work of Bronfenbrenner (1979)—has been widely promoted as an appropriate conceptual framework to shape humanitarian response to children in crisis settings (Boothby, Strang, et al., 2006; Wessells, 2006). It emphasizes how "children develop in a social milieu in which family, peers, teachers and the wider community are part of the fabric of their day-to-day lives" (Kostelny, 2006, p. 25). In this study, this emphasis was most clearly evident in support for the reunification of children with families or other culturally appropriate carers, but was also reflected in encouragement for the prompt reinstitution of schooling (Statement 79) and the prioritization of strengthening "traditional mechanisms" (Statement 23) and wider community activities and institutions.

Although we can identify themes of resilience and social ecological analysis shaping proposed best practice in child care and protection in crisis settings, utilization of such concepts was generally underdeveloped. For example, best practice statements tended to use "protective" factors and "resilience" as a general "shorthand" for strengths that can be capitalized upon. However, useful distinctions have been made in the resilience literature in the way that factors beneficially influence outcomes. *Promotive factors*, for instance, are beneficial in all circumstances, whereas *protective factors* show benefit in circumstances of risk (Masten & Wright, 2009). Luthar et al. (2000) distinguish factors that do not enhance functioning but reduce its deterioration (protection stabilizing), a key function in crisis settings. Factors that may lead to enhanced functioning (so-called flourishing; see Keyes, 2007) on transition into crisis (protection enhancing) have also been the focus of much interest (Luthar et al., 2000; Masten & Obradovic, 2008). These conceptualizations are not a matter of theoretical elegance; they are potentially of major programmatic significance. In the immediate wake of crisis, for instance, factors having a "protection-stabilizing" effect will usually be most important to identify and promote. In subsequent phases of recovery, more directly protective influences might appropriately be fostered.

Recent developments in the field of *development systems theory* also offer potentially valuable insights for humanitarian assistance to children. Lerner (2006) uses the concept of "adaptive developmental regulation" to refer to the two-way interaction between individuals and the many "layers" of their context, and suggests that the aggregate of multiple interactions shapes developmental outcomes. Reviewing four decades of research on resilience, Masten and Wright (2009) conclude the recurrent implication of certain basic human adaptive systems assumed to have been shaped through processes of biological and cultural evolution. Such formulations shift attention from linear "factors" promoting development to nested systems that serve this function. Masten and Obradovic's (2008) grouping of such systems into those promoting "human capital" and "social capital" has certain parallels with formulation of resource domains relevant to psychosocial support in humanitarian settings (Ager, Strang, & Abebe, 2005). This distinction offers encouragement to mobilizing adaptive systems of attachment and agency (reflecting the domain of "human capital"), as well as of family, peer and cultural systems (reflecting the more familiar territory for humanitarian work of "social capital"). The focus on identification, and mobilization, of adaptive systems that support development also has powerful synergies with moves to establish a more systemic approach to child protection work in crisis settings (PFMH, 2008; UNICEF, 2008).

There is a third area where humanitarian practice might benefit better from developmental theory. Explicitly developmentalist conceptions shaped much writing on humanitarian interventions with children in the later decades of the 20th century, principally from a perspective of child psychology (e.g., Garbarino, Kostelny, & Dubrow, 1991; Hundeide, 1991; Jareg, 1987). Although developmental psychology currently has a less prominent role than formerly within developmental science, the latter retains sensitivity to issues of developmental transition. However, there was remarkably little emphasis in the reported statements on such matters. The term *age*, for example, was cited only once (although it was stated as a basis for disaggregation to ensure appropriate programming response; Statement 37). The support given to schooling may be taken as some recognition of the value of this institution to structure children's experience in a developmentally appropriate way, but there was otherwise little explicit attention paid to programming to be shaped to specific "ages and stages." The move from school to work and the value of vocational training was the only developmental transition to which explicit attention was paid (Statements 35, 81, 84, and 91), potentially reflecting the concerns—particularly in conflict settings—of the vulnerability of disengaged youth to recruitment and illegal activity (Stark, Boothby, & Ager, 2009). Beyond the asserted role of restoring stable caregiving, there was no recognition of the specific needs of preschool children, for example, reinforcing other observations calling for a more explicit consciousness of early childhood intervention needs in humanitarian contexts (CGECCD, 2008). Such interventions may be particularly important given the potential for "cascading": early developmental progress providing a crucial foundation for subsequent developmental challenges (Masten & Wright, 2009).

What are the implications for developmental scientists of the findings of the current study? We suggest that there are two major themes that warrant exploration. The first of these is responding to the issue of culture and, specifically, cultural variation. Developmental scientists have increasingly recognized the "culture bound" nature of much contemporary analysis. For example, Weisz, Sigman, Weiss, and Mosk (1993) have argued that data suggest that "core tenets of attachment theory are deeply rooted in mainstream Western thought and require fundamental change when applied to other cultures or minority groups." Rothbaum, Pott, Azuma, Miyake, and Weisz (2000) conclude that "findings from research on parent–child and adult mate relationships suggest that there are different paths of development in Japan and the United States" noting that "the notion that there are different paths of development challenges Western investigators' presumption that certain processes . . . are central in all relationships" (p. 1121). Pena (2007) has noted the challenge of ensuring the linguistic, functional, cultural, and metric equivalence of measures used across contexts. Tudge et al. (2006) have demonstrated the wide cultural variation—and its interaction with class—in the behavior of 3-year-olds attending and not attending day care.

In the face of such evidence, it is not surprising to see humanitarian practitioners reluctant to draw lessons across settings. Participants failed to consensually endorse "necessary protection and care actions are often broadly similar across situations" (Statement 6), adding such comments as "one size never fits all," "approaches should not be cookie cutter," and "globalized standards that bear little relation to local realities are likely to be problematic." Concerted critique over recent years of models that insufficiently recognize the socioeconomic and cultural embeddedness of child well-being (Ahearn et al., 1999; Bracken, Giller, & Summerfield, 1995; Wessells & Monteiro, 2004) has clearly established cultural sensitivity as a dominant lens through which issues of children's development is viewed.

But there is a tension in the approach commended by senior practitioners, as elsewhere there are calls for a stronger "culture of learning" (Statement 14) and more evidence-based advocacy (Statement 40), both requiring some capacity for generalization. There are, as one participant noted, clear challenges with adopting a "blank sheet" approach in settings of acute crisis and vulnerability.

What humanitarians require is a conceptualization of developmental processes that is sufficiently robust to formulate consistent sources of influence on developmental outcomes, without presuming upon their necessary form. In other terms, the practice of humanitarian assistance in diverse cultures requires sensitivity to indigenous psychologies and *emic* perspectives (Kim, Yang, & Hwang, 2006). But it also requires some derived *etic* understandings in order to respond in a timely and strategic manner, drawing global lessons from the experience of, and recovery from, crisis. Providing a basis for such understandings is a great challenge for developmentalists (Rothbaum et al., 2000). But it is a challenge worthy of the aspirations of contemporary developmental science. As Lerner (2006) suggests, "developmental science that is devoid of knowledge of the individual and group ranges among diverse groups, and that is devoid of knowledge of the range of assets in diverse contexts, is an incomplete developmental science" (p. 12). To develop theory and conceptualization fit for application to diverse settings will require engagement with a far broader range of contexts than is currently reflected in the pages of leading journals. Development science must be an increasingly global enterprise, with commitment to research partnerships with "southern" institutions and use of methodologies suited to unstable and resource-poor settings (PFMH, 2008), prerequirements of effective "reach" into the required diversity of development contexts.

Second, theory and evidence needs to be presented in forms that are accessible to policy makers and practitioners, and related to the decisions and actions they face regarding child care and protection in crisis. The strong endorsement of (and frequent reference to) the injunction to "do no harm" (Statement 11) may be welcomed as a recognition of cultural sensitivity and the frequent imbalance of power when working in crisis-affected settings. But it may also be seen as testament to the lack of clear, confident guidance from developmental science regarding key developmental processes (leaving workers, as participants variously noted, "paralyzed" or "with their hands tied from doing even the most basic of psychosocial work").

Luthar and Cicchetti's (2000) 10 "Guiding Principles in Applying the Resilience Perspective Toward Developing Interventions and Policies" provide a good illustration of how developmental science may be presented for practitioner audiences. These principles are of a form not unlike the presented statements here, and with much convergence (e.g., "interventions should target salient vulnerabilities and protective processes that operate across multiple levels of influence" cf. Statement 9; "intervention efforts should aim at fostering services that eventually can become self-sustaining" cf. Statement 23; "interventions must be designed not only to reduce negative influences but also to capitalize on specific resources within particular populations" cf. Statement 25). If developmental science can formulate its findings and implications in such terms, humanitarians are far more likely to reflect such thinking in their programming decisions.

In conclusion, this study suggests significant consensus amongst senior practitioners regarding "best practice" in the care and protection of children in crisis settings. This knowledge base reflects the influence of conceptualizations of social ecology and resilience, but greater use could be made of developmental science in relation to more sophisticated understandings of the manner of "protective" influences, particularly in relation to their being formulated as the reflecting the operation of core human adaptive systems. Some issues that failed to secure consensus support (e.g., regarding expectations of the scalability of interventions and the use of the concept of vulnerability), signal key issues for debate in the field at this time. For developmental scientists to constructively contribute to such debate a more extensive research engagement in diverse cultural settings (including those experiencing crisis) is required, accompanied by clear formulation of guidance for practitioners and policy makers.

References

Ager, A., Boothby, N., & Wessells, M. (2007). The use of consensus methodology in determining key research and practice development questions in the field of intervention with children associated with fighting forces. *Intervention: International Journal of Mental Health, Psychosocial Work and Counselling in Areas of Armed Conflict*, *5*, 124–129.

Ager, A., Strang, A., & Abebe, B. (2005). Conceptualising community development in war-affected populations: Illustrations from Tigray. *Community Development Journal*, *40*, 158–168.

Ahearn, F. (2000). Conclusions and implications for further research. In F. L. Ahearn (Ed.), *Psychosocial wellness of refugees* (pp. 234–238). New York: Berghahn.

Ahearn, F., Loughry, M., & Ager, A. (1999). The experience of refugee children. In A. Ager (Ed.), *Refugees:*

Perspectives on the experience of forced migration (pp. 215–236). London: Continuum.

Annan, J., Blattman, C., & Horton, R. (2006). *The state of youth and youth protection in Northern Uganda: Findings from the Survey for War Affected youth*. Kampala: AVSI and UNICEF.

Beattie, E., & Mackway-Jones, K. (2004). A Delphi study to identify performance indicators for emergency medicine. *Journal of Emergency Medicine, 21*, 47–50.

Berkman, D. S., Lescano, A. G., Gilman, R. H., Lopez, S. L., & Black, M. M. (2002). Effects of stunting, diarrheal disease, and parasitic infection during infancy on cognition in late childhood: A follow-up study. *Lancet, 359*, 564–571.

Betancourt, T., Pochan, S., & de la Soudiere, M. (2005). *Psychosocial adjustment and social reintegration of child ex-soldiers in Sierra Leone*. IRC Final Report.

Boothby, N., Crawford, J., & Halperin, J. (2006). Mozambique child soldier life outcome study: Lessons learned in rehabilitation and reintegration efforts. *Global Public Health, 1*, 87–107.

Boothby, N., Strang, A., & Wessells, M. (2006). *A world turned upside down: Social ecological approaches to children in war zones*. Bloomfield, CT: Kumarian.

Bowling, A. (2002). *Research methods in health: Investigating health and health services*. Buckingham: Open University Press.

Bracken, P., Giller, J., & Summerfield, D. (1995). Psychological responses to war and atrocity: The limitations of current concepts. *Social Science & Medicine, 40*, 1073–1082.

Brainard, L., & LaFleur, V. (2007). *Making poverty history?* Washington, DC: Brookings Institute.

Brazelton, T. E., & Sparrow, J. D. (2006). *Touchpoints 0-3*. Cambridge, MA: Perseus.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brooks, A. (2005). *Disarmament, demobilisation and reintegration of children associated with fighting forces: Lessons learned 1998–2002*. Dakar: UNICEF.

Brown, N., Crawford, I., Carley, S., & Mackway-Jones, K. (2006). A Delphi-based consensus study into planning for biological incidents. *Journal of Public Health, 28*, 238–241.

Bryce, J., & Boschi-Pinto, C. (2005). WHO estimates of the causes of death in children. *Lancet, 365*, 1147–1152.

Buckner, J. C., Mezzacappa, E., & Beardslee, W. R. (2003). Characteristics of resilient youths living in poverty: The role of self-regulatory processes. *Development and Psychopathology, 15*, 139–162.

Cantwell, N. (1997). *Starting from zero: The promotion in children's rights in post-genocide Rwanda*. Florence: International Child Development Centre.

Carballo, M., Heal, B., & Horbaty, G. (2006). Impact of the tsunami on psychosocial health and well-being. *International Review of Psychiatry, 18*, 217–223.

Consultative Group on Early Childhood Care and Development. (2008, October). *Annual Consultation Outcome Report: Thematic focus on early childhood care and development in emergencies*. Consultative Group on Early Childhood Care and Development, Hungary.

Coalition to Stop the Use of Child Soldiers. (2004). *Child soldiers global report 2004*. London: Author.

Cohen, J., Mannarino, A. P., Gibson, L. E., Cozza, S. J., Brymer, M. J., & Murray, L. (2005). Interventions for children and adolescents following disasters. In E. Ritchie, P. Watson, & M. Friedman (Eds.), *Interventions following mass violence and disasters: Strategies for mental health practice* (pp. 227–256). New York: Guilford.

Crawford, I. W. F., Mackway-Jones, K., Russel, D. R., & Carley, S. D. (2004). Planning for chemical incidents by implementing a Delphi based consensus study. *Journal of Emergency Medicine, 21*, 20–23.

Daniel, B., Wassell, S., & Gilligan, R. (1999). *Child development for child care and protection workers*. London: Jessica Kingsley.

Duncan, E. A. S. (2006). The nature and use of consensus methodology in practice. In G. Kielhofner (Ed.), *Research in occupational therapy: Methods of inquiry for enhancing practice* (pp. 401–410). Philadelphia: LA Davis.

Elwyn, G., O'Connor, A., Stacey, D., & the International Patient Decision AIDS Standards. (2006). Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *British Medical Journal, 333*(7565), 417 doi: 10.1136/bmj.38926.629329.AE

Garbarino, J., Kostelny, K., & Dubrow, N. (1991). *No place to be a child: Growing up in a war zone*. Lexington, MA: Lexington Books.

Green, B., Jones, M., Hughes, D., & Williams, A. (1999). Applying the Delphi technique in a study of GPs' information requirements. *Health and Social Care in the Community, 7*, 198–205.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*, 1008–1015.

Hepburn, A. (2006). Running scared: When children become separated in emergencies. In N. Boothby, A. Strang, & M. Wessells (Eds.), *A world turned upside down: Social ecological approaches to children in war zones* (pp. 63–88). Bloomfield, CT: Kumarian.

Hoddinott, J., & Kinsey, B. (2001). Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics, 63*, 409–436.

Hofmann, C.-A., Roberts, L., Shoham, J., & Harvey, P. (2004). *Measuring the impact of humanitarian aid: A review of current practice (HPG Report 17)*. London: Overseas Development Institute.

Hundeide, K. (1991). *Helping disadvantaged children: Psycho-social intervention and aid to disadvantaged children in third world countries*. London: Jessica Kingsley.

Inter-Agency Standing Committee. (2007). *IASC guidelines on mental health and psychosocial support in emergency settings*. Geneva: Author.

Inglehart, R. (1995). Public support for environmental protection: Objective problems and subjective values in 43 societies. *Political Science and Politics*, *28*, 57–72.

Jareg, E. (1987). *Psychosocial factors in relief: Experiences during the Ethiopian famine 1985–86*. Oslo: Redd Barna.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal*, *5*, 311.

Kemp, S., & Burt, C. D. B. (2002). Altruism in valuing government and market-supplied goods. *Journal of Socio-Economics*, *31*, 167–169.

Keyes, C. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist*, *62*, 95–108.

Kim, U., Yang, G., & Hwang, K. (2006). *Indigenous and cultural psychology: Understanding people in context*. New York: Springer.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Kostelny, K. (2006). A culture-based, integrative approach. In N. Boothby, A. Strang, & M. Wessells (Eds.), *A world turned upside down: Social ecological approaches to children in war zones* (pp. 19–38). Bloomfield, CT: Kumarian.

Leaning, J. (2001). Ethics of research in refugee populations. *The Lancet*, *357*, 1432–1433.

Lerner, R. (2006). Developmental science, developmental systems, and contemporary theories of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (Vol. 1, 6th ed., pp. 894–941). New York: Wiley.

Levy-Shiff, R., Hoffman, M. A., & Rosenthal, M. (1993). Innocent bystanders: Young children in war. *Infant Mental Health Journal*, *14*, 116–130.

Luthar, S., & Cicchetti, D. (2000). The construct of resilience: Implications for interventions and social policies. *Development and Psychopathology*, *12*, 857–885.

Luthar, S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*, *71*, 543–562.

Martorell, R., Rivera, J., Kaplowitz, H., & Pollitt, E. (1992). Long-term consequences of growth retardation during early childhood. In M. Hernandez & J. Argente (Eds.), *Human growth: Basic and clinical aspects* (pp. 143–149). Amsterdam: Elsevier Science.

Masten, A. S., & Gewirtz, A. H. (2006). Vulnerability and resilience in early child development. In K. McCartney & D. A. Phillips (Eds.), *Handbook of early childhood development* (pp. 22–43). Malden, MA: Blackwell.

Masten, A. S., & Obradovic, J. (2008). Disaster preparation and recovery: Lessons from research on resilience in human development. *Ecology and Society*, *13*(1), 9. Retrieved March 19, 2010, from http://www.ecologyandsociety.org/vol13/iss1/art9/

Masten, A., & Wright, M. (2009). Resilience over the lifespan: Developmental perspectives on resistance, recovery and transformation. In J. C. Reich (Ed.), *Handbook of adult resilience*. (pp. 213–237). New York: Guilford.

Mendez, M. A., & Adair, L. S. (1999). Severity and timing of stunting in the first two years of life affect performance on cognitive tests in late childhood. *Journal of Nutrition*, *129*, 1555–1562.

Moss, P. A., & Schutz, A. (2001). Educational standards, assessment and the search for consensus. *American Educational Research Journal*, *38*, 37–70.

Murphy, M., Black, N., Lamping, D., McKee, C., Sanderson, C., Askham, J., et al. (1998). Consensus development methods and their use in clinical guideline development. *Health Technology Assessment*, *2*(3), 1–88.

Office for the Coordination of Humanitarian Affairs. (2007). *Protection Cluster Working Group: Mission statement & terms of reference*. Retrieved from http://ocha.unog.ch/humanitarianreform

O'Hara, S. U., & Stagl, S. (2002). Endogenous preferences and sustainable development. *Journal of Socio-Economics*, *31*, 511–527.

Pena, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*, 1255–1264.

Program on Forced Migration and Health. (2008). *Care and protection of children in crisis affected countries: A good practice: Policy change initiative*. New York: Program on Forced Migration & Health, Columbia University. Retrieved May 9, 2010, from http://www.forcedmigration.columbia.edu

Powell, C. (2002). The Delphi technique: Myths and realities. *Methodological Issues in Nursing Research*, *41*, 376–382.

ReliefWeb. (2008). *More than 10,000 children, elderly "alone after quake."* Retrieved June 12, 2008, from http://reliefweb.int/rw/RWB.NSF/db900SID/RMOI-7EYKC5?OpenDocument

Rothbaum, F., Pott, M., Azuma, H., Miyake, K., & Weisz, J. (2000). The development of close relationships in Japan and the US: Pathways of symbiotic harmony and generative tension. *Child Development*, *71*, 1121–1142.

Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, *28*, 763–774.

Schon, D. (1983). *How professionals think in action*. New York: Basic Books.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction*. Thousand Oaks, CA: Sage.

The Sphere Project. (2004). *Humanitarian charter and minimum standards in disaster response*. Retrieved May 19, 2010, from http://www.sphereproject.org

Stark, L., Boothby, N., & Ager, A. (2009). The reintegration of children associated with fighting forces: Ten years on from Cape Town. *Disasters*, *33*, 522–547.

Toole, M. J., & Waldmen, R. J. (1997). The public health aspect of complex emergencies and refugee situations. *Annual Review of Public Health*, *18*, 283–312.

Tudge, J. R. H., Doucet, F., Odero, D., Sperb, T. M., Piccinini, C. A., & Lopes, R. S. (2006). A window into different cultural worlds: Young children's everyday activities in the United States, Brazil, and Kenya. *Child Development, 77,* 1446–1469.

UNAIDS. (2004). *Children on the brink: A joint report of new orphan estimates and a framework for action.* New York: UNAIDS/UNICEF/USAID.

United Nations High Commissioner for Human Rights. (1989). *The convention on the rights of the child.* UN General Assembly Resolution 44/25, October 1989, Geneva: Author.

United Nations International Children's Emergency Fund. (2008). *Humanitarian action report 2008.* New York: UNICEF. Retrieved from http://www.unicef.org/har08/index_index.html

Walker, A. M., & Selfe, J. (1996). The Delphi method: A useful tool for the allied health researcher. *International Journal of Therapy and Rehabilitation, 3,* 677–681.

Walker, S. P., Chang, S. M., Powell, C. A., & Grantham-McGregor, S. M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *Lancet, 366,* 1804–1807.

Walker, S. P., Wachs, T., Gardner, J., Lozoff, B., Wasserman, G., Pollitt, E., et al. (2007). Child development: Risk factors for adverse outcomes in developing countries. Lancet Series 2: Child development in developing countries. *Lancet, 369,* 145–157.

Weinstein, J. (2005). *Disentangling the determinants of successful demobilization and reintegration.* Center for Global Development, Working Paper No. 69.

Weisz, J. R., Sigman, M., Weiss, B., & Mosk, J. (1993). Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. *Child Development, 64,* 98–109.

Wessells, M. (2006). *Child soldiers: From violence to protection.* Cambridge, MA: Harvard University Press.

Wessells, M., & Monteiro, C. (2004). Psychosocial assistance to internally displaced people in Angola: A child focused, community-based approach. In K. Miller & L. Rasco (Eds.), *Refugee mental health: Ecological approaches to adaptation and recovery* (pp. 67–94). Mahwah, NJ: Erlbaum.

Wessells, M., & Strang, A. (2006). Religion as resource and risk. In N. Boothby, A. Strang, & M. Wessells (Eds.), *A world turned upside down: Social ecological approaches to children in war zones* (pp. 199–222). Bloomfield, CT: Kumarian.

Williams, J. R. A., Hyder, T., & Nicoli, S. (2005). Save the children's experience: ECD in emergencies. Early childhood matters. *Responses to Young Children in Post-Emergency Situations, 124.*

Williamson, J., & Cripe, L. (2002). *Assessment of DCOF supported child demobilization and reintegration activities in Sierra Leone.* Washington, DC: USAID.

Wood, A., Apthorpe, R., & Borton, J. (2001). *Evaluating international humanitarian action: Reflections from practitioners.* London: Zed Books.

World Health Organization. (2004). *The importance of caregiver–child interaction for the survival and healthy development of young children: A review.* Geneva: Author.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1287–1303

Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations

Erik D. Thiessen
Carnegie Mellon University

Meagan N. Yee
Indiana University

Whereas young children accept words that differ by only a single phoneme as equivalent labels for novel objects, older children do not (J. F. Werker, C. J. Fennell, K. M. Corcoran, & C. L. Stager, 2002). In these experiments, 106 children were exposed to a training regime that has previously been found to facilitate children's use of phonemic contrasts (E. D. Thiessen, 2007). The results indicate that the effect of this training is limited to contexts that are highly similar to children's initial experience with the phonemic contrast, suggesting that early word-form representations are not composed of entirely abstract units such as phonemes or features. Instead, these results are consistent with theories suggesting that children's early word-form representations retain contextual and perceptual features associated with children's prior experience with words.

From early in life, children are able to remember words; they recognize words after familiarization (e.g., Jusczyk & Aslin, 1995) and look at the appropriate visual referent of spoken words (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Tincofff & Jusczyk, 1999). But speech presents a representational challenge because of its variability. The same word can be produced with different rates, fundamental frequencies, amplitude, and formant structure both within and across speakers. Learners must represent words in a sufficiently flexible manner that two different utterances of the same word can be mapped to the same meaning. But at the same time, representations must be conservative enough that learners can determine when they are hearing a novel word (as when a child who knows *dog* hears *bog* for the first time), as opposed to a novel production of a familiar word. This is likely to be especially challenging for infants and young children, whose knowledge of the phonological system of their native language is in a state of flux (e.g., Werker & Tees, 1984).

One approach to this representational problem suggests that speech is mapped on to abstract, invariant representations (e.g., Nearey, 1997). For example, in traditional linguistic theory (e.g., Chomsky & Halle, 1968) representations of word forms in the mental lexicon are constructed from abstract, invariant phonological components such as phonemes or features (e.g., Nearey, 2001; Prince & Smolensky, 1997). These components are identical across different surface realizations: the same features are activated by the [d] in *dog* and the [d] in *lad*. According to this approach, listeners must factor out the surface variability of speech to access the abstract units that compose the word. There have been several proposals as to how this might be accomplished, including perceptual normalization accounts (e.g., Shankweiler, Strange, & Verbrugge, 1977) and articulatory theories (e.g., Browman & Goldstein, 1986, 1992; Repp, 1982). These proposals, and others, share the assumption that surface variability, such as speaker-specific indexical properties, is noise that must be discarded before listeners can identify abstract, phonological units of representation.

An alternative approach suggests that the surface variability of speech is not discarded but instead remains part of the long-term representation of word forms from early in development (e.g., Werker & Curtin, 2005). This approach is consistent

---

This research was funded by a grant from the National Science Foundation (BCS-0642415) to the first author. We thank Ashley Episcopo, Teresa Pegors, and many other research assistants for their help in running participants. Additionally, we thank Lori Holt, Katharine Graf Estes, and Janet Werker and two anonymous referees for helpful discussion and useful comments on previous versions of this manuscript, as well as the many parents who were willing to participate in the research.

Correspondence concerning this article should be addressed to Erik D. Thiessen, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. Electronic mail may be sent to thiessen@andrew.cmu.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0018

with demonstrations that listeners recognize recurring words more easily when they are produced by the same speaker, indicating that indexical information is encoded in representations (e.g., Bradlow, Nygaard, & Pisoni, 1999; Goldinger, 1996). One way listeners might encode surface variability is to store multiple exemplars of words rather than a single abstract and canonical version (e.g., Goldinger, 1998; Jusczyk, 1993). Such information may be sufficient for word recognition alone, or it may be incorporated in parallel with more abstract representations (e.g., Hanson, 1977; Remez, Fellowes, & Rubin, 1997). Even very young learners may have access to both kinds of representations. Research on categorical perception (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) and vowel change detection in the face of speaker differences (Kuhl, 1983) suggest that infants can respond to acoustic differences that reflect linguistically relevant distinctions, such as phonemic categories, in the face of surface variability. At the same time, however, infants appear to encode speaker-specific properties of their experience with words (Houston & Jusczyk, 2003), and respond to acoustic differences between exemplars of a single phonemic category (McMurray & Aslin, 2005).

In the adult state, an abstract approach and approaches that incorporate surface variability in representations make similar predictions. For example, connectionist models can learn, through exposure to variable input, to respond to the same phonemic distinctions as models relying on more abstract representations (e.g., Vallabha, McClelland, Pons, Werker, & Amano, 2007). However, these two kinds of representational approaches lead to very different predictions about the rate of learning and generalization. In a theory where generalization occurs automatically at the level of abstract phonemes, a child who responds to [d]—as distinguished from [t]—has mapped the spoken realization of those phonemes onto representations abstracted from their acoustic idiosyncracies. Because these representations are abstract, they should generalize to a variety of novel acoustic instances. A learner who responds to the contrast followed by the vowel [i] should respond to the contrast in the same way when it is followed by the vowel [a]; even though the acoustic realization of [d] is different in *da* than *di*, the underlying representation is identical. Similarly, a contrast available in syllable-initial position should be available in syllable-medial or syllable-final position, insofar as the contrast can be mapped onto identical representations.

These kinds of abstract approaches predict very rapid patterns of developmental acquisition. Shvachkin (1948/1973)—working from an acoustic framework—argued that once children master the use of a particular phoneme, it should be available in every context in which the native language employs that phoneme. More recent theories have made similar predictions. For example, Lindblom, Diehl, Park, and Salvi (2008), whose work emphasizes a more articulatory approach, predict that "**once a target** [place feature] **has been learned in one context, it can be immediately re-used in other contexts**." (bolding in original). This powerful generalization is one of the strengths of an abstract representational system (e.g., Berent, Marcus, Shimron, & Gafos, 2002), as it can potentially explain the rapidity, and consistency across a variety of differing environments, with which children acquire language (for discussion, see Lindblom, Diehl, Park, & Salvi, in press). By contrast, a representational system that encodes multiple aspects of the surface variability of a contrast should lead to less inclusive generalization in early learning. Because the acoustic realization of a phonemic contrast is different across different word positions and vocalic contexts, a contrast that is productive in one setting may not be productive in all settings. Instead, generalization on this kind of account should be based, at least in part, upon the similarity of the current instance to prior experience (e.g., McClelland & Elman, 1986). Further, the frequency with which a learner experiences a contrast in a particular context should be important; if the [d]–[t] contrast is more common in word-initial position than in word-final position (e.g., Zamuner, 2009), a learner might initially respond to contrast more successfully in word-initial position (e.g., Schafer & Mareschal, 2001; Storkel, 2001).

Note, however, that theories proposing an abstract representational system need not predict invariant responses to every instance of a phoneme. For example, in word recognition, different realizations of the phoneme may be more or less effective in activating the phonemic representation, as is the case when an ambiguous articulation is heard (e.g., Norris & McQueen, 2008). Similarly, in phonotactic learning, a particular phoneme may occur in some word positions or syllabic contexts, but not in others (e.g., Hayes, 2004). More generally, it is well known that distinctions at a perceptual level need not correspond with distinctions at a phonological level (e.g., Dietrich, Swingley, & Werker, 2007; Pallier, Colome, & Sebastian-Galles, 2001; Stager & Werker, 1997). Indeed, even in early childhood, there is evidence that children's sensitivity is graded as a function of the tasks in which they are tested. Children will use a distinction in a word-recognition task—as when distinguishing between familiar words like *ball* and *doll* (e.g., Fennell & Werker, 2003; Swingley & Aslin, 2000)—that they fail to use in the context of associating a novel label with a novel object (Pater, Stager, & Werker, 2004). These complexities make it particularly important to explore generalization in the context of word–object association. Simply because infants generalize in a perceptual task (e.g., Maye, Weiss, & Aslin, 2008) does not necessarily indicate that they will do so when representing word forms. Conversely, the fact that learners generalize conservatively when learning phonotactic regularities (e.g., Prince & Tesar, 2004) does not require that they will do so when learning words.

To distinguish theories where generalization occurs at the level of abstract phonemes, or context-free features (e.g., Dinnsen, O'Connor, & Gierut, 2001; Nearey, 2001), from perceptually based accounts (e.g., Houston & Jusczyk, 2003; Ryalls & Pisoni, 1997; Singh, 2008), it is critical to assess the course of learning in a task relevant to lexical development. Learning should be rapid and highly generalizable if children's representations are abstract, slower and more conservative for representations that incorporate subphonemic and indexical information (cf. McClelland & Patterson, 2002). One important component of word learning is the ability to associate labels with objects, an ability that can be assessed experimentally (e.g., Stager & Werker, 1997). Such assessments indicate that young children have difficulty differentiating between novel labels that differ by only a single phoneme (such as *bih* and *dih*), although they can learn to do so successfully (Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009). In this series of experiments, we will assess how widely children generalize their learning about phonemic differences in a word–object association task.

Using the difference between words that differ by a single phoneme (i.e., minimal pair words such as *bog* and *dog*) in a word–object association task is difficult for children in the first half of their 2nd year of life (e.g., Shvachkin, 1948/1973; Stager & Werker, 1997). Upon learning a label for a novel object (such as *bin*), children of around 14 months of age accept a wide variety of minimal pairs of that word (e.g., *pin* and *din*) as labels for the novel object (Pater et al., 2004). This is not due to an inability to hear the phonemic distinction. When tested on the same contrasts with known words (such as *ball* and *doll*), children are responsive to the identical contrast (e.g., Fennell & Werker, 2003; Swingley & Aslin, 2002). Instead, children appear to have difficulty using phonemic contrasts in the context of associating novel labels with novel objects. The ability to use phonemic contrasts to differentiate novel labels improves with development. By 17 months, children are beginning to behave more like adults: After they have learned a label for a novel object, they reject alternative labels that differ on one or more phonemes. At 14 and 17 months, the ability to do so is related to vocabulary size; by 20 months, this is no longer the case, perhaps because by this age virtually all children have exceeded the threshold of vocabulary necessary to succeed in the task (Werker et al., 2002).

The relation between vocabulary and performance on the switch task posited by Werker et al. (2002) suggests that linguistic experience influences children's use of phonemic contrasts in when learning words. This would explain why children with larger vocabularies were more successful in the task. Consistent with this hypothesis, computational modeling indicates that experience with words may affect the way that children respond to phonemic contrasts (Schafer & Mareschal, 2001), and both 17- and 20-month-olds know far more words than 14-month-olds (Fenson et al., 2002). In particular, children in this age range know relatively few minimal pair words (e.g., Caselli et al., 1995; Charles-Luce & Luce, 1995). For example, Swingley and Aslin (2007) found that, in a sample of 18-month-old Dutch-learning children, over two thirds of the words in children's vocabularies had no minimal pair neighbors. The majority of the words with which children become familiar between 14 and 20 months provide examples of phonemes distributed in different lexical contexts: [d] in *doggy* and *daddy*, [t] in *teddy* and *tiger*.

Thiessen (2007) demonstrated that exposure to phonemes in different lexical contexts ([d] in *dawbow* and [t] in *tawgoo*) facilitates children's use of the phonemic contrast in a word–object association task. This may be related to the phenomenon of acquired distinctiveness: When two similar stimuli, A and B, are paired with two highly discriminable associates X and Y (forming compound stimuli AX and BY), the distinction between A and B is heightened (e.g., Hall, 1991). Children's familiarity with an increasing number of words in the 2nd year of life, very few of which are minimal pairs, may provide evidence that different phonemes are paired with different lexical forms in the language and facilitate use of the distinction between them

(Thiessen, 2007). This would explain the observed correlation between vocabulary size and use of phonemes in word–object association tasks (Werker et al., 2002); children with larger vocabularies have more evidence about the distribution of phonemes in their language. From the perspective of exemplar-based theories, these results can be explained in terms of variability among exemplars facilitating generalization (e.g., Hintzman, 1986; Rost & MacMurray, 2009). From the perspective of theories emphasizing more abstract representational systems, or parallel encoding of acoustic and abstract information (e.g., Remez et al., 1997), success on the task might indicate a shift to more abstract encoding. For example, Werker and Curtin (2005) suggest that while children's representations contain rich acoustic detail, children must weight the phonemic aspects of these representations more heavily to succeed in word-learning tasks relying on a single phonemic distinction.

As such, while the results of Thiessen (2007) indicated that exposure to phonemes in different lexical contexts improves children's abilities to use the contrast between phonemes, those experiments were ambiguous with respect to whether children learned about abstract, readily generalized contrasts (i.e., phonemes or features). The current series of experiments systematically assessed how widely infants generalize a contrast (such as [d] vs. [t]) after they have been exposed to stimuli that should facilitate their use of the contrast in a word–object association task. If children encode speech abstractly, generalization should be rapid and pervasive: Once a child has identified the [d]–[t] contrast in one context, it should be available in many (or even all) contexts (e.g., Lindblom et al., 2008; Shvachkin, 1948/1973). By contrast, if children encode subphonemic or indexical information, generalization may be much more limited, especially early in learning (e.g., McClelland & Patterson, 2002). Children should initially be most likely to use the [d]–[t] contrast in a context they have previously experienced, and less likely to use it in contexts that differ from the original context. Conversely, training in a different context should allow children to succeed in contexts where they have previously failed, as is demonstrated in Experiment 3.

Experiment 1

Fifteen-month-olds fail to respond to differences in minimal pair labels in a word–object association task (e.g., Pater et al., 2004; Shvachkin, 1948/1973). Thiessen (2007) found that exposing children to words in which phonemes occurred in clearly distinct lexical contexts—as [d] and [t] in *dawbow* and *tawgoo*—facilitates their use of the distinction between minimal pairs. However, these results are inconclusive with respect to the degree of specificity or abstraction of children's learning. Children were tested with the syllables *daw* and *taw* after exposure to the words *dawbow* and *tawgoo*. Therefore, children could potentially have learned about a feature contrast (voicing), a phonemic contrast ([d]–[t]), or a contrast in a specific syllabic context (*daw* vs. *taw*). If children learned about voicing, this knowledge should generalize to a wide variety of contexts. By contrast, if infants learned about the specific syllables, then their generalization should be limited. Similarly, most prior experiments testing children's use of phonemic contrasts expose participants to phonemes in only one particular context, making it difficult to assess how generalizable children's representation of these contrasts might be (e.g., Stager & Werker, 1997).

Our goal in this experiment was to assess how widely children generalize the contrasts they use in word–object association tasks. To do so, we exposed children to the same training regime used in Thiessen (2007): Children learned the names of three objects. The labels of two of these objects provided evidence of phonemes ([d] and [t]) occurring in distinct lexical contexts: *dawbow* and *tawgoo*. The third object was labeled *yad*. After learning the names of these three objects, infants were presented with test trials in which the third object was paired with its original label (*yad*), or a novel minimal pair label (*yat*). Experience with *dawbow* and *tawgoo* enables children to differentiate between *daw* and *taw* test trials in this procedure (Thiessen, 2007). If children have learned about a phonemic contrast, or a featural contrast (voicing), this learning should generalize to a novel word position (from [d]-[t]aw to ya[d]-[t]). By contrast, if children's learning is more specific, and tied to the particular exemplars with which children have prior experience, then exposure to *dawbow* and *tawgoo* would not be as informative with respect to *yad* as it is with respect to *daw*, and children may fail to use the contrast between [d] and [t] in word-final position.

Method

Participants

Participants were 16 toddlers (half female; 13 Caucasian, 2 Asian, and 1 African American) between the ages of 14.5 and 15.5 months

(M = 14.77). To obtain data from these 16 infants, it was necessary to test 24. The additional 8 participants were excluded for the following reasons: fussing or crying (7), experimental error (1). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems. Children were recruited from the Pittsburgh area via mailing. The participants primarily came from middle- or upper-middle-class socioeconomic status (SES) families.

Stimuli

Participants were exposed to three novel objects, each paired with a unique label. The three objects were identical to those used by Stager and Werker (1997) and Thiessen (2007). Each object was animated and moved against a black background. Each of the three labels was produced by the SoftVoice computer synthesizer, set to a female voice, and a monotonic production with a fundamental frequency of 220 Hz.

The first object was paired with the label *yad* [jaed] or the label *yat* [jaet] (here and elsewhere, brackets indicate IPA transcription). Each infant heard only one of these labels—either *yad* or *yat*—during the habituation phase. The second object was paired with the label *dawbow* [dabo], and the third object was paired with the label *tawgoo* [tagu]. The labels *dawbow* and *tawgoo* were the identical synthesized recordings as those used in Thiessen (2007), and the labels *yad* and *yat* were synthesized to have matching pitch, amplitude, and vocal quality. Following Thiessen's original experiment, there was only a single token of each synthesized label used during the habituation and test phases. Each label was repeated for as long as the object remained on screen, with pauses of 1.4 s between repetitions.

Procedure

This experiment used a variant of the habituation procedure employed by Stager and Werker (1997) procedure, and identical to that used in Thiessen (2007). Participants, seated on a parent's lap in a sound-attenuated room, controlled the duration of the presentation of the stimuli by the length of their gaze at a central monitor. To eliminate bias, parents wore headphones, and the experimenters sat in an adjacent room, blind to the nature of the stimulus being presented. The experimenter coded the duration of the child's looking time online, using a Macintosh running the Habit program for OS X (Cohen, Atkison, & Chaput, 2004). After the child reached the habituation criterion (looking time less than 50% of the average of the first three trials), six test trials were presented.

The child controlled the duration of stimulus presentation, during both the habituation and test phases, by gazing at a 32′ video monitor 150 cm in front of their seated position. The child's attention was attracted to the monitor by a colorful video of Winnie the Pooh coupled with a recorded verbal encouragement. Once the experimenter determined the child's attention was fixated on the monitor, stimulus presentation was initiated. An object then appeared on the screen and the speakers adjacent to the monitor began to repeat the label associated with that object. The stimulus presentation continued until the child looked away for more than 1.5 s, or until the child had gazed at the monitor for 20 s (the maximum time allowed per trial). The video of Winnie the Pooh appeared at the end of each trial to recapture the child's attention.

The experiment began with two familiarization trials, in which participants saw a nonsense object paired with the word *neem* (a recording from a trained female speaker). The familiarization trials were used to help participants become accustomed to the pairing of audio and visual stimuli. Once these trials were finished, the habituation phase began.

During the habituation trials, one of the three novel objects appeared on screen, and was paired with a unique label (Object 1 was always paired with *yad* or *yat*; Object 2 was always paired with *dawbow*, and Object 3 was always paired with *tawgoo*). The objects were presented in blocks of three trials, and the order of presentation was random within those blocks. As in Thiessen (2007), each participant was required to complete at least two habituation blocks, or their data were excluded (all participants reached this threshold). Each object moved onscreen while the label was repeated with pauses of 1.4 s between each repetition. The object and the label were presented until the child looked away from the monitor for 1.5 s, at which point the attention-getting stimulus reappeared. Looking times to each trial were calculated in real time, and the habituation trials continued in random order until the child met the habituation criterion: Average looking time for three consecutive trials that fell below 50% of their looking time to the first three habituation trials.

Once the child met the habituation criterion, the six test trials began. In all test trials, the child saw Object 1 on the monitor. There were two kinds of test trials: same and switch trials. In the same trials, the children heard the syllable that they had previously heard paired with Object 1. For half of the infants, this was *yad*; for the other half of the infants, this was *yat*. In the switch trials, Object 1 was paired with a minimal pair of the label infants had heard in the habituation phase. For infants exposed to the object paired with *yad*, the switch trial paired the object with the novel label *yat*. The opposite was true for infants exposed to *yat* during the habituation phase.

Same and switch trials alternated, and the nature of the initial test trial was counterbalanced across participants. As in the habituation trials, the object stayed on the screen, and the label continued to repeat, for as long as the participant continued to look at the monitor.

Results and Discussion

On average, children habituated in 8.9 trials. Although children habituated in fewer trials to *yad* ($M = 7.5$) than to *yat* ($M = 10.2$) during the habituation phase, the difference was only marginally significant, $t(14) = 1.9$, $p = .08$, Cohen's $d = 0.55$. More importantly, children's preference for same versus switch test trials did not differ as a function of which item they heard during the habituation phase, $t(14) = 1.5$, $p = .15$, Cohen's $d = 0.41$. Therefore, for all subsequent analyses, participants were grouped together, regardless of which item they heard during the habituation phase. During the habituation phase, infants spent an average of 96.5 s ($SD = 28.6$) looking to the objects during the habituation phase. Thus, on average infants accumulated approximately 30 s of looking to the object labeled *yad/yat*.

Our primary question was whether infants in the test phase responded differentially to trials in which the novel object was paired with its original label (same trials), versus trials in which the object was paired with a minimal pair of that label (switch trials). As illustrated in Figure 1, children did not respond differentially to same and switch trials. A two-tailed t test (all t tests reported are two-tailed) indicated that the difference in looking times between same and switch trials was not significant, $t(15) < 1$, *ns*, Cohen's $d = 0.04$.

These results contrast sharply with those of Thiessen (2007). In that experiment, exposure to *dawbow* and *tawgoo* lead children to respond differentially to objects paired with *daw* versus *taw*. Here, after exposure to *dawbow* and *tawgoo*, children responded equivalently *yad* and *yat*. One possibility is that the sample size in this experiment was too small to yield a significant result; the sample size in Thiessen was 24 infants, compared to only 16 in this experiment. A power analysis indicated that for an effect the size in the Thiessen paper (approximated by a Cohen's d of .75, given the reported difference of 1.9 s between groups and a standard deviation of 2.4 s), the probability of identifying a significant result with the current sample size of 16 is only 54%. A sample size of 30 would yield a greater than 80% probability of finding an effect. Thus, in Experiment 2 we utilized a larger sample size. Another possibility is that children are unable to hear the difference between [d] and [t] in word-final position. However, the results of Experiment 3 (below) indicate that children can respond to this contrast in word-final position.

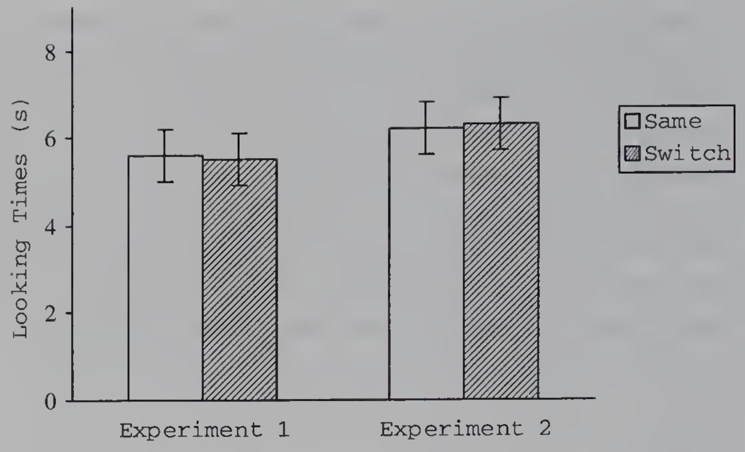


Figure 1. Children's looking time to same and switch trials after exposure to *Dawbow* and *Tawgoo*.
Note. In Experiment 1, same and switch trials are *yad* versus *yat*; in Experiment 2, they are *dee* versus *tee*. Error bars indicate ± standard error.

There is, though, a theoretical possibility consistent with the lack of a significant difference found in this experiment. What children learn from experiencing speech sounds in distinct lexical contexts (as with the [d] and [t] sounds in *dawbow* and *tawgoo*) does not lead, at least initially, to generalizations about abstract units such as features or phonemes. Instead, children may be learning about more specific, contextually bound units. That is, children's learning may be limited to situations that share a high degree of similarity to their initial experience. For example, exposure to *dawbow* and *tawgoo* may only facilitate use of [d] and [t] in word-initial position, or even be limited to a particular syllabic context. Whereas children fail to use the [d]–[t] contrast in word-final position, they may succeed if tested with contexts in which the [d]–[t] contrast is realized in a way that is more similar to its occurrence in *dawbow* and *tawgoo*.

Experiment 2

Writing and phonemic systems use the same symbols to indicate a phoneme in different word positions. However, the acoustic realization of phonemes differs in as a function of context. For example, the phoneme [t] is aspirated in *tear*, while it is unaspirated in *steer*. The [t] in *steer* sounds, when the preceding fricative is removed, much like a [d] (e.g., Maye, Werker, & Gerken, 2002). Similarly, phonemes are articulated quite differently in word-initial position than in word-final position, and these articulatory differences lead to differences in the acoustic realization of the phoneme (Stevens, 1998). If children are representing speech sounds in a nonabstract manner, the acoustic differences between word-initial and word-final position may make it difficult for children to generalize from information about a phonemic contrast across word positions; experience with *dawbow* and *tawgoo* may not be readily applicable to *yad* and *yat*.

However, testing generalization by presenting phonemes in word-final position may be especially challenging for young children. The ends of words appear to be less salient than word-onset positions for children (e.g., Clark, 1991), and word-initial information may be especially salient given its role in word recognition (e.g., Marslen-Wilson, 1987; Rodd, 2004). Given children's potential difficulty with word-final position, asking children to generalize from word-initial to word-final position may be an overly conservative test of generalization. In Experiment 2, we tested generalization from one word-initial setting to a second word-initial setting, in a different syllabic context. As in Experiment 1, children were exposed to three novel object–label pairings: objects labeled *dawbow*, *tawgoo*, and *dee*. According to Thiessen (2007), experience with *dawbow* and *tawgoo* facilitates children's use of the contrast between *daw* and *taw*. If children are learning about a relatively abstract contrast, they should also be facilitated in their use of the contrast between *dee* and *tee*.

Method

Participants

Participants were 30 toddlers (half female; 26 Caucasian and 4 Asian) between the ages of 14.5 and 15.5 months ($M = 15.22$). To obtain data from these 30 children, it was necessary to test 37. The additional 7 participants were excluded for the following reasons: fussing or crying (5), experimental error (2). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems. Children were recruited from the Pittsburgh area via mailing. The participants primarily came from middle- or upper-middle-class SES families.

Stimuli

The computer-animated objects were identical to those used in Experiment 1, as were the words *dawbow* and *tawgoo*. The only difference between the stimuli used in Experiment 1, and the stimuli used in this experiment, was that Object 1 was paired with the labels *dee* [di] or *tee* [ti] (each infant heard only one during the habituation phase). These labels replaced *yad* and *yat* from Experiment 1, and were synthesized using the same acoustic parameters as the stimuli in that experiment. As in Experiment 1, there was only a single repeated token of each label.

Procedure

The procedure was identical to that of Experiment 1. After the habituation phase, infants were presented with same and switch trials. On same trials, Object 1 was paired with the same label as it was during the habituation phase. On switch trials, it was paired with a minimal pair of that label. Thus, for infants who heard *dee* during the habituation phase, the switch trial presented the word *tee*.

Results and Discussion

On average, children exposed to *dee* (in addition to *dawbow* and *tawgoo*) habituated in 13.4 trials, while children exposed to *tee* (in addition to *dawbow* and *tawgoo*) habituated in 6.9 trials. This difference was significant, $t(28) = 2.7$, $p < .05$, Cohen's $d = 1.02$. Infants exposed to *dee* spent an average of 128.1 s ($SD = 75.4$) looking to the objects during the habituation phase. Infants exposed to *tee* spent an average of 79.4 s ($SD = 35.4$) looking to the objects during the habituation phase. This difference was also significant, $t(28) = 2.3$, $p < .05$, Cohen's $d = 0.83$. It is unclear why children found one set of habituation trials more interesting than the other. It may be indicative of a preference for the speech synthesizer's production of voiced items over voiceless items, or a difference in children's familiarity with these particular syllables in their natural environment. However, any possible item preference did not influence children's performance on the test items: There was no significant difference in children's preference for same versus switch trials as a function of their exposure to *dee* or *tee*, $t(28) < 1$, *ns*, Cohen's $d = 0.06$.

Given that the differences in rate of habituation suggested that children may have performed differently as a function of exposure, we analyzed the group data separately, using a 2 (habituation: *dee* or *tee*) × 2 (test trial: same vs. switch) analysis of variance (ANOVA). There was no main effect of habituation, $F(1, 28) < 1$, *n.s.*, $\eta_p^2 = .01$. Nor was there any interaction between habituation and test item, $F(1, 28) < 1$, *ns*, $\eta_p^2 = .01$.

Most importantly, there was no difference in children's looking time to same or switch trials $F(1, 28) < 1$, *n.s.*, $\eta_p^2 = .01$. Figure 1 illustrates the average looking time to same and switch trials, averaged across both groups of infants (habituated to *dee*, or habituated to *tee*).

These results suggest that exposure to *dawbow* and *tawgoo* had little, if any, effect on children's use of the [d] and [t] contrast in a novel syllabic context. Whereas *dawbow* and *tawgoo* facilitated children's use of the contrast in the same (*daw–taw*) syllabic context (Thiessen, 2007), children continued to treat minimal pair words differentiated by the contrast interchangeably, when the contrast was embedded in a novel syllable. This indicates that what children may not readily generalize at the level of abstract phonemes, or context-free features. Instead, children's learning appears to be strongly tied to specific, previously experienced, contexts.

Experiment 3

The results of Experiment 1 and 2, taken together, suggest that children are not generalizing at the level of abstract units such as phonemes or features. Rather, their use of speech sounds to indicate contrasts in meaning appears to be contextually bound. If so, then it should be easiest to facilitate children's use of a phonemic contrast in a particular context by providing them distributional evidence about that phoneme in an identical context. To test this hypothesis, children in Experiment 3 were tested using the *yad–yat* pairing from Experiment 1. However, children received distributional information about the [d]–[t] contrast in word-final position (the words *boeyad* vs. *gooyat*). This experience with the contrast in word-final position should facilitate its use in *yad–yat*, unlike the experience with the phonemes in word-initial position in Experiment 1.

To increase the ecological validity of the comparison between word-initial and word-final position, Experiment 3 used natural speech, as opposed to synthesized speech; this necessitated a replication of the data from Experiment 1 with the new, natural stimuli. As such, there were three conditions in this experiment.

Single Object

In this condition, children were habituated to a single object, paired with the label *yad*. Following the results of Stager and Werker (1997), children should fail to respond differentially to same (*yad*) and switch (*yat*) trials.

Word Initial

This condition replicates Experiment 1. In addition to *yad*, children were habituated to two objects labeled *dawbow* and *tawgoo*, providing information about the distribution of [d] and [t] in word-initial condition. If the results of natural speech parallel the results with synthesized speech, children should also fail to respond differentially to same and switch trials.

Word Final

In this condition, children were habituated to objects labeled *yad*, *boeyad* [bojaed], and *gooyat* [gujaet]. This condition provides evidence that the [d]–[t] contrast occurs in different distributions in word-final position, and thus should be the condition in which children are most successful at responding differentially to the contrast in word-final position.

Method

Participants

Participants were 60 toddlers (half female; 51 Caucasian, 4 Asian, 4 African American, and 1 Hispanic) between the ages of 15.0 and 16.5 months ($M = 15.4$). Twenty were assigned to the single-object condition, 20 were assigned to the word-initial condition, and 20 were assigned to the word-final condition. To obtain data from these 60 participants, it was necessary to test 73. The additional 13 infants were excluded for the following reasons: fussing or crying (8), experimental error (2), parental interference (2), and looking times averaging < 2 s during test trials (1). According to parental report, all children were free of ear infection at the time of testing, and reported no history of hearing problems. Children were recruited from the Pittsburgh area via mailing. The participants primarily came from middle- or upper-middle-class SES families.

Stimuli

The computer-animated objects were identical to those used in Experiment 1. However, the audio stimuli were recorded from utterances produced by a female native speaker of English, and digitally edited to match the amplitude of the stimuli used in Experiments 1 and 2. The test items (*yad* and *yat*) were produced by the same speaker. As in prior experiments, only a single token of each label was used during the habituation and test phases.

Procedure

All participants saw Object 1 paired with the nonsense word *yad*. In the single-object condition, this was the only word–object pairing to which children were habituated. In the word-initial condition, children saw two additional objects, labeled *dawbow* and *tawgoo*. This replicates Experiment 1, providing distributional information about the [d]–[t] contrast in word-initial position.

The third condition was the word-final condition, in which children were provided with distributional information about the [d]–[t] contrast in word-final position. In this condition, children were habituated to Object 1 paired with *yad*, and two additional objects paired with *boeyad* and *gooyat*.

The habituation procedure used was identical to that of Experiment 1 and 2. After habituation, all infants were tested on same and switch trials, as in Experiment 1. On same trials, Object 1 was paired with the familiar label *yad;* on switch trials, it was paired with the novel label *yat*.

Results and Discussion

Participants in the single-word condition habituated in fewer trials ($M = 6.8$) than participants in either the word-initial ($M = 10.2$) or word-final condition ($M = 9.8$). Similarly, participants in the single-word condition spent less overall time looking during the habituation phase ($M = 58.3$ s, $SD = 28.5$) than infants in either the word-initial ($M = 112.1$ s, $SD = 47.6$) or word-final condition (104.6, $SD = 39.1$). This is unsurprising, as participants in the word-initial and word-final conditions were exposed to three word-object pairings during the habituation phase, whereas participants in the single-object condition were only exposed to one. These results mimic those of Thiessen (2007), where children habituated more quickly to a single object than multiple objects.

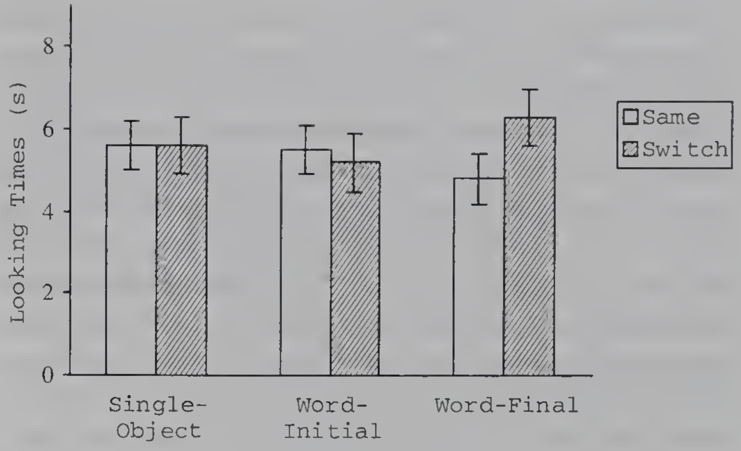


Figure 2. Children's looking time to same and switch trials in the three conditions of Experiment 3.
Note. The difference between same and switch trials was significant only in the word-final condition. Error bars indicate ± standard error.

As seen in Figure 2, participants in the single-object condition responded equivalently to same and switch trials, $t(19) < 1$, *ns*, Cohen's $d = 0.01$. Children's lack of differentiation between object labels differing by only a single phoneme replicates Stager and Werker's (1997) original experiment, and many other demonstrations that children at this age have difficulty responding to single-phoneme differences in word–object association tasks (e.g., Pater et al., 2004; Shvachkin, 1948/1973; Werker et al., 2002).

In the word-initial condition (in which infants were exposed to *dawbow* and *tawgoo*, in addition to *yad*), children also failed to respond differentially to same and switch trials, $t(19) < 1$, *ns*, Cohen's $d = 0.14$. Children's looking times to same and switch trials are illustrated in Figure 2; their failure to respond differentially to the test trials replicates Experiment 1.

Most importantly, as seen in Figure 2, children in the word-final condition responded differentially to same and switch trials. This difference was significant, $t(19) = 2.4$, $p < .05$, Cohen's $d = 0.6$. Unlike experience with *dawbow* and *tawgoo* (the word-initial condition), exposure to *boeyad* and *gooyat* facilitated children's response to the difference between *yad* and *yat*.

To more thoroughly explore the difference between the word-initial and word-final condition, we performed a 2 (condition) × 2 (test trial type) ANOVA. There was no main effect of condition, $F(1, 38) < 1$, *ns*, $\eta_p^2 = .01$. This suggests that looking times were equivalent across the two conditions. There was also no main effect of test item, $F(1, 38) = 1.2$, $p = .27$, $\eta_p^2 = .03$. This lack of effect is not surprising, as participants did not demonstrate a consistent preference for items across conditions. Most importantly, however, there was an interaction between condition and item, $F(1, 38) = 5.1$, $p < .05$, $\eta_p^2 = .12$. This interaction indicates that toddlers' preference in the word-final condition significantly differs from their lack of preference in the word-initial prediction. These results support the hypothesis that children's performance is better when they receive distributional information in the same context as the phonemes are presented in the test trials.

The difference between the word-initial and word-final conditions is informative about the nature of children's early generalization of speech sounds. In both conditions, children received distributional evidence that the phonemes [d] and [t] occur in different lexical forms; this kind of distributional evidence has previously been found to facilitate children's use of phonemic contrasts (Thiessen, 2007). However, only distributional information in the same syllable and word position facilitated children's ability to use the contrast at test. Children in the word-initial condition, who received distributional information about the phonemes in different syllables and word positions, performed like children in the single-object condition who received no distributional information at all. This suggests that children's generalization is limited, at least initially, to contexts that are highly similar to their prior experience.

One potential confound needs to be addressed: There is a possibility that children respond differently in the two conditions, not because of the context of the distributional information they received but because they receive different amounts of exposure to the stimuli. Morgan, Duran, and Layton (2005) have suggested that the ability to dishabituate in experiments of this type is mediated by amount of exposure; when children are overexposed to a particular sound, they become less sensitive to it. Alternatively, children may have received less exposure to the *yad*-object in the word-initial condition and thus not learned the pairings between sound and object. To assess these possibilities, we performed a two-tailed t test comparing length of exposure during the habituation phase for children in the word-initial condition ($M = 112.1$ s, $SD = 47.6$) and children in the word-final condition (104.6, $SD = 39.1$). There was no significant difference, $t(38) < 1$, *ns*, Cohen's $d = 0.17$. This suggests that length of exposure does not explain why children differentiate between same and switch trials in the word-final condition, and fail to do so in the word-initial condition. Instead, this analysis reinforces the hypothesis that the critical difference in these conditions relates to the similarity between the distributional information children receive during testing and the context in which phonemic distinctions occur during test trials.

As such, these results are inconsistent with accounts in which children's early representations of speech are organized in such a way as to allow automatic generalization at the level of abstract phonemes. If phonemic representations were entirely abstract (e.g., bundles of features), children's use of those phonemes should generalize widely (e.g., Berent et al., 2002; Lindblom et al., 2008; Shvachkin, 1948/1973). Instead, the current results suggest that children are storing some amount of contextual (i.e., not abstracted) detail about their prior experience with phonemes. This detail may be related to word position, syllabic context, or some other acoustic aspect of the specific realization of the previously experienced phoneme. These details persist and lead to performance that is better when the context matches training than when it does not. While the precise nature of children's representation is still ambiguous (see the General Discussion), the fact that children perform better in the word-final condition than in the word-initial condition indicates that children's representations of early word forms are rich with contextual detail.

General Discussion

To master their native language, children must learn which sound contrasts indicate a difference in meaning and which do not. While children make remarkable progress in that regard during their 1st year of life (e.g., Werker & Tees, 1984), they still have some difficulty using these distinctions in a word–object association task into the 2nd year (e.g., Stager & Werker, 1997). By 17 months, children are more successful at making use of at least some of these single phoneme differences (Werker et al., 2002), although they may not be proficient with every contrast present in their native language until later (Shvachkin, 1948/1973). While a number of explanations have been proposed for this developmental difference—including increases in capacity (Stager & Werker, 1997) and understanding of referential intent (e.g., Fennell, 2008)—the current results provide support for accounts suggesting that older children are successful because they have more information about the distribution of phonemes in their native language (e.g., Thiessen, 2007).

Children's vocabularies grow dramatically during the 2nd year of life (e.g., Fenson et al., 2002). During this time, most of the words children learn provide evidence of phonemes occurring in different lexical contexts (like /b/ and [d] in *baby* and *diaper*). Young children know comparatively few words in which two different phonemes occur in identical contexts (minimal pair words like *ball* and *doll*), far fewer, both absolutely and as a relative proportion of their vocabulary, than adults (e.g., Coady & Aslin, 2003). As demonstrated by Experiment 3, and also by Thiessen (2007), experiencing phonemes in different lexical contexts helps children to use the phonemic distinction in word–object association tasks. Further, children with larger vocabularies, and thus more evidence that phonemes are distributed in different words, are more successful using phonemic distinctions in word–object association tasks (Werker et al., 2002). These results converge to suggest that vocabulary growth plays an important role in children's use of phonemic contrasts by providing children with distributional information.

The current results are novel in that they indicate that children's generalization of the distributional information they receive is somewhat limited. Children are more likely to take advantage of distributional information about a phonemic contrast when that contrast is tested in the same context as the original distributional information. While these results do not rule out the possibility that some aspects of children's representation of word forms are abstract (e.g., featural), they indicate that children must be storing some contextual detail about their prior experience with phonemes. This is inconsistent with theories that argue that representation of speech is primarily or completely abstract. For example, most featural theories of representation argue that phonemes can be reduced to bundles of abstract features that are invariant across word position or syllabic context (e.g., Chomsky & Halle, 1968; Cole & Scott, 1974; Eulitz & Lahiri, 2004). It is a common assumption that children's representations of speech can be characterized by the same kinds of abstract representations (e.g., Dinnsen et al., 2001). The context specificity of the current results is inconsistent with abstract representations like features or phonemes, although it may be the case that abstract representations and contextual acoustic detail are processed separately and in parallel (e.g., Hanson, 1977; Remez et al., 1997). One possibility is that abstract representations, such as phonemes, can incorporate contextual detail such as the syllabic context in which the phoneme occurred in ways that limit the generalizations learners make about those phonemes in other contexts (e.g., Prince & Tesar, 2004).

Indeed, these results, and others like them (e.g., Houston & Jusczyk, 2003) are consistent with accounts positing that representations of speech are not composed of abstract, phonologically pure components (e.g., Goldinger, 1996). While phonemes and features are useful descriptively, the evidence for the psychological reality of phonemes has been criticized on a number of grounds (e.g., Lotto & Holt, 2003). Many phenomena once thought to be decisive evidence in favor of abstract representations, such as context effects, have now been demonstrated to occur in the absence of phonemic knowledge (e.g., Lotto, Kluender, & Holt, 1997; Mann, 1986). Similarly, intraphonemic variation, which should be discarded on abstract representational accounts, clearly plays a role in adult speech perception (e.g., Volaitis & Miller, 1992). Although both adults and infants appear to respond to phonological categories in ways that are consistent with abstract representations (e.g., Eimas et al., 1971), it may be the case that nonabstract systems could accomplish the same feats (Bybee & McClelland, 2005). For example, when infants generalize on the basis of a feature such as voicing (e.g., Maye et al., 2008), this generalization may be due to acoustic similarities among voiced consonants, rather than a commonality in the abstract featural representation + voicing. However, caution is due in interpreting the current results; evidence that children's representations are not completely abstract is not logically sufficient to rule out some degree of abstraction.

Some authors have suggested that the degree of abstraction in children's representation varies by the task in which they are engaged, or their developmental state (e.g., Shvachkin, 1948/1973). Child phonologists have proposed a discontinuity between the types of representations used in perceptual tasks (such as speech discrimination) and the representations used for word learning and language more broadly (e.g., Barton, 1980; Brown & Matthews, 1997; Ferguson & Farwell, 1979). One possibility that this raises is that children's representations in the current tasks looked relatively concrete because they were not treating the task as truly linguistic, but rather as a perceptual association task. On this kind of account, had the training been embedded in a more naturalistic setting, children may have engaged a different set of representations and generalized their phonemic usage more widely. By contrast, Werker et al. (2002) have proposed that the underlying representations used by infants in both perceptual and linguistic tasks are identical. Due to the fact that word-learning and perceptual tasks typically differ in their difficulty (word-learning tasks are more difficult because they require forming an association between words and objects), children show differing degrees of sensitivity to phonemic variation in the two kinds of tasks. While the current data do not differentiate between these two accounts, other work indicates that lessening the difficulty of word-learning tasks—whether by using familiar objects (e.g., Fennell, 2006), or prefamiliarized labels (e.g., Swingley, 2007)—leads infants to perform more successfully in word–object association tasks. This pattern of results is more consistent with theories positing a continuous representational system than theories positing discontinuous representational systems.

This should not be taken as evidence that there are no developmental differences in children's representations over time. A number of theorists have argued that more abstract representations emerge over time (e.g., Best, Tyler, Gooding, Orlando, & Quann, 2009; Dietrich et al., 2007). Werker and Curtin (2005), for example, have proposed that children's representational system encodes subphonemic and indexical information throughout the course of development. As children gain more experience with their native language, they come to attend more to phonemic aspects of these representations than to more idiosyncratic factors. This may explain why older children are more successful in using phonemic detail in word–object association tasks (e.g., Werker et al., 2002). The emergence of less context-specific phonemic distinctions may occur as a function of children's experience with these contrasts in more varied contexts (e.g., Rost & MacMurray, 2009). In these experiments, children failed to generalize a phonemic contrast after experiencing it in a single vocalic or syllabic context. However, generalization may be possible after experiencing the contrast in multiple contexts, as suggested by models of generalization in memory (e.g., Hintzman, 1986). As a contrast is experienced in multiple different contexts, the information that is consistent—the contrast itself—is emphasized, while the variable contextual information is less supported in memory. The emergence of behavior governed by (seemingly) abstract units from contextually detailed representations is consistent with a variety of formal models.

However, before the predictions of such models can be fully evaluated, it is critical to understand the nature of the contextual detail encoded in children's representations. While the current results demonstrate that children's representations of speech incorporate some kind of contextual detail about their prior experience with phonemes, the nature of that contextual information is still uncertain. Two (potentially complementary) possibilities have been suggested in prior theories of linguistic development: context at a grain size above the phoneme, such as syllables, and more granular acoustic context, such as coarticulatory information. The syllable is a universal feature of human linguistic systems; additionally, syllables are both perceptually salient and relatively limited in number. Infants appear to be sensitive to the presence and identity of syllables in speech (e.g., Bertoncini, Bijeljic-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987). Because of these facts, a number of theorists have argued that syllables may play an important role in children's early representations of speech (e.g., Mehler, Dupoux, & Segui, 1990). On these kinds of accounts, the syllable is an undifferentiated unit, not composed of combinatorially distinct subcomponents. This lack of subcomponents (phonemes or features) renders a syllabic account consistent with the current results. When children are exposed to *dawbow* and *tawgoo*, they are learning about the syllables *daw* and *taw*, not the phonemes [d] and [t] or the features [−voice] and [+voice]. Therefore, there is little reason to expect generalization to syllable contrasts like *yad/yat* or *dee/tee*. This syllable-level account is consistent with theories of phonotactic learning in which generalization is conditioned upon the syllabic context in which a particular phoneme has been previously encountered (e.g., Prince & Tesar, 2004).

An alternative, although not mutually exclusive, possibility is that different instances of particular phonemes might be differentiated by fine-grained acoustic differences in the realization of the phoneme itself. Both infants and adults are sensitive to intraphonemic differences: differences in the articulation of two exemplars of the same phonemic category (e.g., McMurray & Aslin, 2005; Volaitis &

Miller, 1992). Representation of these intraphonemic differences is often assumed in exemplar memory models of speech representation (e.g., Goldinger, 1998). Similarly, the WRAPSA model of speech perception (Jusczyk, 1993, 1997) proposes that the developmentally earliest form of representation is a spectral and temporal code derived immediately from the acoustic signal itself. The fact that infants recognize words more easily after a delay when they are spoken by the same speaker than when they are produced by a novel speaker (Houston & Jusczyk, 2003; Singh, 2008) is consistent with approaches emphasizing a fine-grained acoustic basis of representation, because it indicates that infants' representation of the word forms preserves idiosyncratic acoustic indexical information such as rate of speech. If young children are representing intraphonemic acoustic details, generalization should also be more limited than is the case with an abstract phonological system. This is because different speech sounds are realized quite differently in different contexts, due to coarticulatory processes. The realization of [d] in *yad* is quite different than the realization in *dawbow*. By contrast, the [d] in *boeyad* is nearly identical to the [d] in *yad*, making distributional information from *boeyad* much easier to apply to *yad* than distributional information from *dawbow*.

Subsequent experimentation will be necessary to assess the precise nature of the contextual detail present in young children's representations of word forms. The current results demonstrate that children fail to generalize to a segment that shares a feature (like voicing) in a very different syllabic position, or followed by a very different vowel. But children might be successful if they were tested on the same feature—in the same syllabic and vocalic context—at a different place of articulation. For example, children might successfully use the distinction between *baw* and *paw* after exposure to *dawbow* and *tawgoo* (we thank an anonymous reviewer for this suggestion). Alternatively, children may generalize to a different syllable as long as the vocalic context and syllabic position are the same. That is, after exposure to *dawbow* and *tawgoo*, children may use the distinction between *dock* and *talk*. If children are representing intraphonemic acoustic variation, they should be successful in this training scenario (a variant of which we are currently running in our lab), since the phonemes [d] and [t] are articulated in the same way in *dawbow* and *dock* and *tawgoo* and *talk*, occurring in the same syllabic position and vocalic context. By contrast, if children are representing syllabically, they may fail in this training regime, since *dock* and *talk* are novel syllables.

The current results do not distinguish between a syllabic and coarticulatory account of the contextual information stored in young children's representations of early word forms (for a discussion, see Jusczyk, 1993; Coady & Aslin, 2004). But on either account, consistent with the current results, children's representations are not characterized solely by abstract phonemic or featural components. This leads to a different view of the developmental processes between the age when children fail to use phonemic differences in word-learning tasks (Stager & Werker, 1997), and the age when they succeed (Werker et al., 2002). If children are representing words using featural or phonemic representations, the change between failure and success is somewhat mysterious, as children are already using appropriate units. One way to resolve this dilemma is to propose that children's early representations of speech are holistic and underspecified (e.g., Walley, 1993). For example, a word like *dog* might, early in lexical development, be represented only by the single feature +alveolar. However, recent research on comprehension is largely inconsistent with an underspecification account (although see Kager, Van der Feest, Fikkert, Kerkhoff, and Zamuner, 2007, for production data consistent with underspecification). Young children are highly sensitive to even slight mispronunciations of known words (e.g., Swingley & Aslin, 2000, 2002, 2007), and even some novel words (Mani & Plunkett, 2008).

Rather than underspecification, these results suggest that young children's early word learning may fail to take advantage of phonemic differences due to the richness of the input and children's representations (cf. Werker & Curtin, 2005). There are a multitude of acoustic differences to which children are sensitive (e.g., Houston & Jusczyk, 2003; McMurray & Aslin, 2005). Some of these acoustic differences indicate differences in meaning, but many do not. Children may need to receive information from their linguistic environment about which of these differences are informative in word-learning tasks before they begin to use them productively. For example, hearing examples of two different phonemes distributed in different word forms helps children subsequently use that phonemic contrast when using novel words (Thiessen, 2007). The current results are consistent with accounts suggesting that children's representations are rich and multidimensional, insofar as they indicate children's representations are not entirely abstracted but instead incorporate contextual detail. To the extent that children's representations incorporate variability, children may need to receive evidence about which facets of that variability are productive in particular tasks.

This kind of contextual sensitivity, coupled with a learning account, leads to novel predictions about the developmental time course of children's use of different phonemic contrasts. Abstract representational accounts assert that the development in children's use of phonemic contrasts should be sudden (e.g., Lindblom et al., 2008; Nazzi & Bertoncini, 2003); once a child has learned to use a featural contrast (e.g., voicing) in one situation, that contrast can be widely used. By contrast, the current results indicate limited generalization and predict a more graded development in children's use of phonemic contrasts in word learning. More common contrasts should begin to be used earlier than less common contrasts. Similarly, a sound that occurs disproportionately often in word-initial position should be productive in that position earlier than it is productive in other word positions (cf. Schafer & Mareschal, 2001). While there is currently little data to assess these predictions (although see Shvachkin, 1948/1973; Thiessen & Yee, 2008), the current results provide an empirical demonstration that children's use of a phonemic contrast in one context does not immediately generalize to other contexts.

Note, however, that the present findings indicating conservative generalization do not mean that young learners are incapable of generalizing their experience with speech sounds. In particular, prior research suggests that variability in the input facilitates generalization (e.g., Gomez, 2002; Lively, Logan, & Pisoni, 1993). Maye et al. (2008) found that infants who learned to discriminate a featural contrast were able to generalize that learning to a novel contrast varying on the same feature; in that experiment, infants were exposed to multiple exemplars of the featural contrast. Participants in these experiments received only a single exemplar of each to-be-learned label. Young children may well be able to generalize in word–object association tasks if they receive training that more robustly supports generalization beyond specific contexts, such as experiencing a phonemic contrast (e.g., [d]–[t]) in multiple syllabic contexts. This would be consistent with both prior work (e.g., Maye et al., 2002; Rost & MacMurray, 2009) and the current experiments, indicating that infants are sensitive to the distribution of acoustic details in the input and that this distribution plays an important role in the development of children's linguistic abilities.

References

Barton, D. (1980). Phonemic perception in children. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Phology: Vol. 2 perception* (pp. 97–116). New York: Academic.

Berent, I., Marcus, G. F., Shimron, J., & Gafos, A. I. (2002). The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition*, *83*, 113–139.

Bertoncini, J., Bijeljic-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representation of speech sounds. *Journal of Experimental Psychology: General*, *117*, 21–33.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, *20*, 539–542.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, *61*, 206–219.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewen & J. Anderson (Eds.), *Phonology yearbook* (pp. 219–252). Cambridge, UK: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 67–112). Amsterdam: Benjamins.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, *22*, 381–410.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*, 159–199.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*, 727–735.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clark, E. V. (1991). Acquisitional principles in lexical development. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 31–71). Cambridge, UK: Cambridge University Press.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*, 441–469.

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, *89*, 183–213.

Cohen, L. B., Atkison, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in* *infant perception and cognition studies (Version 1.0)*. Austin: University of Texas.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81*, 348–374.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 454–464.

Dinnsen, D. A., O'Connor, K. M., & Gierut, J. A. (2001). The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition. *Journal of Linguistics, 37*, 503–525.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*, 577–583.

Fennell, C. T. (2006, June). *Object familiarity affects 14-month-old infants' use of phonetic detail in novel words*. Paper presented at the biennial International Conference on Infant Studies, Kyoto, Japan.

Fennell, C. T. (2008, March). *Does coarticulatory information aid use of phonetic detail at 14 months?* Paper presented at the International Conference on Infant Studies.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*, 245–264.

Fenson, F., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (2002). *MacArthur Communicative Development Inventories: User's guide and technical manual*. Baltimore: Paul H. Brookes.

Ferguson, C. A., & Farwell, C. B. (1979). Words and sounds in early language acquisition. *Language, 51*, 419–439.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228–231.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Hall, G. (1991). *Perceptual and associative learning*. Oxford, UK: Clarendon.

Hanson, V. L. (1977). Within-category discriminations in speech perception. *Perception and Psychophysics, 21*, 423–430.

Hayes, B. (2004). Phonological acquisition in Optimality Theory: The early stages. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition* (pp. 158–203). Cambridge, UK: Cambridge University Press.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1143–1154.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21*, 3–28.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1–23.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*, 648–654.

Kager, R., Van der Feest, S. V. H., Fikkert, P., Kerkhoff, A. O., & Zamuner, T. S. (2007). Representations of [voice]: Evidence from acquisition. In J. van der Weijer & E. J. van der Torre (Eds.), *Voicing in Dutch* (pp. 41–80). Amsterdam: John Benjamins.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263–285.

Lindblom, B., Diehl, R., Park, S., & Salvi, G. (2008, June). *(Re)use of place features in voiced stop systems: Role of phonetic constraints*. Paper presented at the annual FONETIK conference, Gothenburg, Sweden.

Lindblom, B., Diehl, R., Park, S., & Salvi, G. (in press). Sound systems are shaped by their users: The recombination of phonetic substance. In N. Clements & R. Ridouane (Eds.), *Where do features come from? The nature and sources of phonological primitives*. Elsevier: North-Holland.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

Lotto, A. J., & Holt, L. L. (2003). The illusion of the phoneme. In S. J. Billings, J. P. Boyle, & A. M. Griffith (Eds.), *Chicago Linguistic Society: Vol. 35. The panels* (pp. 191–204). Chicago: Chicago Linguistic Society.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America, 102*, 1134–1140.

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science, 11*, 53–59.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition, 24*, 169–196.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71–102.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Science, 6*, 465–472.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*, B15–B26.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 236–262). Cambridge, MA: MIT Press.

Morgan, J., Duran, P., & Layton, D. (2005, November). *A U-shaped pattern in infants' lexical representations?* Paper presented at the annual Boston University Conference on Language Development.

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*, 136–142.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America, 101*, 381–394.

Nearey, T. M. (2001). Phoneme-like units in speech perception. *Language and Cognitive Processes, 16*, 673–681.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Pallier, C., Colome, A., & Sebastian-Galles, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science, 12*, 445–449.

Pater, J., Stager, C. L., & Werker, J. F. (2004). The lexical acquisition of phonological contrasts. *Language, 80*, 361–379.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to Universal Grammar. *Science, 275*, 1604–1610.

Prince, A., & Tesar, B. (2004). Learning phonotactic distributions. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition* (pp. 245–291). Cambridge, UK: Cambridge University Press.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651–666.

Repp, B. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. *Psychological Bulletin, 92*, 81–110.

Rodd, J. M. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in word recognitions. *Psychonomic Bulletin & Review, 11*, 434–439.

Rost, G., & MacMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*, 339–349.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 3*, 441–452.

Schafer, G., & Mareschal, D. (2001). Modeling infant speech sound discrimination using simple associative networks. *Infancy, 2*, 7–28.

Shankweiler, D., Strange, W., & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Shvachkin, N. Kh. (1973). The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 91–127). New York: Holt, Rinehart, & Winston. (Original work published 1948)

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*, 833–870.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321–1337.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology, 43*, 454–464.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*, 99–132.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language, 56*, 16–34.

Thiessen, E. D., & Yee, M. N. (2008, March). *Evidence that variability in lexical frames affects children's use of phonemic contrasts.* Paper presented at the International Conference on Infant Studies.

Tincofff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*, 172–175.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of* *the National Academy of Sciences of the United States of America*, *104*, 13273–13278.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*, 723–735.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286–350.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*, 1–30.

Werker, J. F., & Tees, R. C. (1984). Cross language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Yoshida, K., Fennell, C., Swingley, D., & Werker, J. F. (2009). 14-month-olds learn similar-sounding words. *Developmental Science*, *12*, 412–418.

Zamuner, T. S. (2009). The structure and nature of phonological neighbourhoods in children's early lexicons. *Journal of Child Language*, *36*, 3–21.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1304–1319

An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children

Petra H. J. M. Vlamings and
Lisa M. Jonkman
Maastricht University

Chantal Kemner
*Maastricht University and University
Medical Center Utrecht*

There is converging evidence for the presence of a fast subcortical face-processing route that operates on global face characteristics in the mature brain. Until now, little has been known about the development of such a route, which is surprising given suggestions that this fast subcortical face-processing route might be affected in neurodevelopmental disorders such as autism. To address this, early visual event-related potentials to pictures of fearful and neutral faces containing detailed or global information in 3- to 4-year-old ($n = 20$), 5- to 6-year-old ($n = 25$), and 7- to 8-year-old ($n = 25$) children were compared. In children, emotional processing was driven by detailed information. Developmental effects are discussed in terms of maturation of the fast subcortical face-processing route as well as an increase in experience with facial expressions with age.

Decoding the emotional content of a face is an important skill in daily life because it helps to evaluate the state and the intentions of others. Given that facial expressions like fear, anger, or threat may be signals of potential danger, it is plausible that the detection and processing of emotional expression proceeds very fast. Several authors have investigated what information the visual system extracts to decode emotional expressions so rapidly (Holmes, Winston, & Eimer, 2005; Pourtois, Dan, Grandjean, Sander, & Vuilleumier, 2005; Vuilleumier, Armony, Driver, & Dolan, 2003; Winston, Vuilleumier, & Dolan, 2003). Visual information that enters our visual system consists of luminance variations, which occur in a wide range of so-called spatial frequencies and can be expressed in cycles per degree (cpd) of visual angle (e.g., De Valois & De Valois, 1988; Goldstein, 1999). Lower spatial frequencies (LSF) represent the large-scale variations, that is, coarse visual information such as shading or large contours, as well as the global configuration of the face (e.g., distance between eyes, eyebrows and mouth). High spatial frequencies (HSF) represent tighter gradients of luminance changes, that is, fine visual information (sharp edges, contours of features such as the mouth, eyes, wrinkles, and so on; Goffaux & Rossion, 2006; Morrison & Schyns, 2001).

By varying the spatial frequency content of a stimulus or image (by filtering out either high or low frequency information), one can study the difference in processing of this information (e.g., De Valois & De Valois, 1988; Goldstein, 1999).

Behavioral studies have provided evidence for differential sensitivity to HSF and LSF content of emotional expressions for face recognition or identification in adults. For instance, it has been shown that participants use LSF information to categorize rapidly emotional expressions, whereas HSF information seems to be most important for the success with which one can explicitly rate emotional expressiveness (or intensity), a process that may proceed at a slower time-scale (Schyns & Oliva, 1999; Vuilleumier et al., 2003; Deruelle & Fagot, 2005; however, see Goren and Wilson, 2006, for different effects using synthetic faces). In addition, the LSF components of faces are critical to the production of rapid behavioral or attentional responses toward fearful facial expressions (Holmes, Green, & Vuilleumier, 2005).

The authors gratefully thank Sabine Koenraads, Pia Jansen, Anne Hermans, and Eveline Eerdekens for helping with the data collection.

Correspondence concerning this article should be addressed to Petra H. J. M. Vlamings, Cognitive Neuroscience Department, Developmental Cognitive Neuroscience Section, Faculty of Psychology and Neuroscience, University of Maastricht, PO Box 616, 6200 MD Maastricht, Netherlands. Electronic mail may be sent to p.vlamings@psychology.unimaas.nl.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0019

Several authors have hypothesized that the rapid processing of fear occurs through a subcortical route in the brain that rapidly processes fear based on LSF information (Pourtois et al., 2005; Vuilleumier et al., 2003; Winston et al., 2003). This route starts at the ganglion cells in the retina. There are two cell types in the retina that are differentially sensitive to HSF and LSF, respectively: parvocellular (small receptive fields) and magnocellular neurons (large receptive fields). Several studies have indicated that although parvocellular and magnocellular neurons respond to an overlapping range of SFs, parvocellular neurons are preferentially tuned (most active) to HSF, whereas magnocellular neurons are more sensitive to LSF information (Enroth-Cugell & Robson, 1966). Because the conduction velocity of axons of the parvocellular system is lower compared to the magnocellular system (see Hess, 2004), signals from the parvocellular system are transmitted more slowly compared to the magnocellular system (see for review Skottun & Skoyles, 2007). The amygdala, a key structure for emotion processing, receives direct input from subcortical regions (tectum and pulvinar) that primarily receive retinal magnocellular input (Pourtois et al., 2005; Vuilleumier et al., 2003; Winston et al., 2003). It has been proposed that this type of subcortical input would enable the amygdala to rapidly process facial expressions based on LSF and to provide quick signals about for example fear-related stimuli (Johnson, 2005; Pourtois et al., 2005; Vuilleumier et al., 2003; Winston et al., 2003).

Two functional magnetic resonance imaging (fMRI) studies indeed provided evidence for this and found that the amygdala and connected subcortical regions are primarily sensitive to low spatial frequency information: Enhanced activity to emotional expressions in the amygdala and other subcortical structures was found only for faces containing LSF, but not HSF, information (Vuilleumier et al., 2003; Winston et al., 2003). Similar effects were found in the fusiform gyrus, a cortical structure specifically involved in the visual processing of faces (Vuilleumier et al., 2003; Winston et al., 2003). Furthermore, two event-related brain potential (ERP) studies in adults found that rapid effects of emotion in visual regions were primarily mediated by LSF information (Pourtois et al., 2005; Vlamings, Goffaux, & Kemner, 2009). Based on these findings, a rapid, subcortical route (tecto-pulvinar-amygdala pathway) for facial expression processing has been proposed that operates on LSF information and modulates activity in visual cortical areas in which emotional signals are further processed (e.g., Le Doux, 1996; Vuilleumier et al., 2003; see for review Johnson, 2005). This fast subcortical face-processing route is seen as a "quick and dirty" route that extracts emotion based on global face characteristics or low spatial frequencies but not detailed information or HSF, which are processed through the "slower cortical route" (Johnson, 2005; Pourtois et al., 2005; Vuilleumier et al., 2003; Winston et al., 2003). Although the above-mentioned studies solely tested the role of SF in fear-related processing, this does not exclude the possibility that other types of facial expressions (e.g., threat-related) might also be processed by this subcortical face-processing route making use of LSF information. In addition, it is not known to what extent this (LSF) route distinguishes between different facial expressions (e.g., fear and surprise).

In sum, behavioral as well as neuroscientific studies in adults have provided evidence that the rapid processing of facial expressions is primarily driven by LSF information and that this might be driven by a rapid subcortical face-processing route that is preferentially tuned to LSF information. This quick and dirty route is thought to bypass the slower cortical route and is thought to modulate directly activity of specialized cortical areas in response to faces.

Little is known about the role of LSF and HSF in the perception of facial expressions in children. As described previously, there is converging evidence from studies in adults for a subcortical face-processing route that is rapid, operates on LSFs, and modulates cortical processing. The development of this pathway has not been studied, but Johnson (2005) suggested that newborn face processing and face preferences might primarily rely on this subcortical route or a precursor of it. This suggestion is based on evidence that face processing in newborns is primarily driven by subcortical pathways or brain structures as the visual and other cortical areas are still relatively immature.

In addition, the infant's limited visual capacities play a role in face processing in the 1st year of life. The vision of a newborn is solely based on LSF information (Acerra, Burnod, & de Schonen, 2002; Morison & Slater, 1985; Slater & Sykes, 1977). This is caused by immaturity of the visual system at several levels, such as lower photoreceptor density and shorter segment length at the retina level, relatively less developed parvocellular neurons (compared to magno-) at the lateral geniculate nucleus level and lower synaptic density and larger cortical receptive field size in the visual cortex (for reviews, see Ellemberg, Lewis, Liu, & Maurer, 1999; Hammarrenger et al., 2003). In agreement with this, a recent study by de Heering et al. (2008) found that face recognition in newborns is primarily steered by very low SF (< 0.5 cpd) information.

However, during infancy and early childhood, the visual system develops and other cortical areas also mature. Vision rapidly increases and, by the age of 3–4 years, sensitivity to HSF as measured by the contrast sensitivity function reaches maturity (e.g., Adams & Courage, 2002). The contrast sensitivity function measures the amount of contrast that is needed to detect a stimulus with a certain SF. Until now, no studies have addressed whether the type of information, HSF or LSF, that is important for rapid facial expression processing changes during development as children become able to perceive a larger range of spatial frequencies.

There is only one behavioral study (Deruelle & Fagot, 2005) that reported that 5- to 8-year-old children, *like adults* (Schyns & Oliva, 1999; Vuilleumier et al., 2003), rely on HSF information when explicitly processing emotional expressions from faces (deciding whether a face is smiling or not). There is, however, no information on whether the *rapid* processing of facial expressions in children is, as in adults, based on LSF information.

A better understanding of normal development of facial expression processing is important as facial expressions play an important role in the social-emotional development of children. In this study, we investigate the role of SF in facial expression processing in children 3–8 years old using ERPs. We primarily investigate the early stages of face processing in the visual areas of the brain, where interactions between SF and emotion have been found to occur in previous studies (Pourtois et al., 2005; Vlamings et al., 2009). Two visual peaks were studied: P1 and N170. The P1 is a positivity peak that occurs 100 ms after stimulus onset (SO). It is a fast exogenous response that reflects striate as well as extrastriate visual processing (e.g., Rossion et al., 1999). The N170 (negativity at 170 ms after SO) originates from a network of regions, probably including the fusiform gyrus, inferior occipital cortex, superior temporal sulcus and the inferior, middle and superior temporal gyri (Henson et al., 2003). The N170 is the earliest "face-specific" ERP component as its amplitude has been found to be consistently enhanced in response to faces in comparison to multiple other nonface, object categories (see for review Jacques & Rossion, 2004). A larger or faster P1 has been found for fearful compared to neutral expressions in adults (Eimer & Holmes, 2002; Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Pourtois et al., 2005; Williams et al., 2004) as well as in infants (starting from 7 months of age) and children (Batty & Taylor, 2006; Dawson, Webb, Carver, Panagiotides, & McPartland, 2004; Nelson & de Haan, 1996). Other studies report effects of emotional expressions on the N170 peak. Typically, larger N170 amplitudes are seen in response to (negative) emotional expressions than to neutral or positive expressions in adults (Batty & Taylor, 2003, 2006; Blau, Maurer, Tottenham, & McCandliss, 2007; Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; Dawson et al., 2004; de Haan, Nelson, Gunnar, & Tout, 1998; Stekelenburg & de Gelder, 2004). In children, effects of (negative) emotional expression on the N170 have been reported in 3- to 4-year-olds (Dawson et al., 2004) and 5-year-olds (de Haan et al., 1998). However, a recent study by Batty and Taylor (2006) investigating facial expression processing from early childhood to adolescence indicated that early effects of negative emotion do not occur at the face-specific N170 until the age of 14–15. The latencies of these early effects at both P1 and N170 are usually somewhat delayed in infants and children compared to adults. Although the P1 and N170 results are not entirely consistent across developmental studies, the above-mentioned ERP studies underline the rapid (< 200 ms) processing of emotional expressions in visual areas in both adults and children.

This study investigated effects of SF manipulations on the early processing of fearful facial expressions in the visual cortex in groups of 3- to 4-year-old, 5- to 6-year-old, and 7- to 8-year-old children using the event-related potential technique. During a passive viewing task, we examined whether the effects of emotion expression at early visual information processing stages (P1 and N170) are primarily mediated by global LSF cues, as in adults, or by detailed HSF information. Through administration of an additional active behavioral task in which participants were required to categorize rapidly negative (fearful) or neutral face expressions, we aimed to explore the development of the role of spatial frequency on emotion recognition at the behavioral level. That is, we tested whether a LSF or HSF advantage for the processing of emotion would be reflected in reaction times. This behavioral task was only administered to the 5- to 6-year-old and 7- to 8-year-old children.

Method

Participants

This study included 20 children (11 female, 9 male) between 3 and 4 years (mean age = 3.10 years), 25 children (12 female, 13 male) between 5 and 6 years (mean age = 6.1 years), and 25 children (12 female, 13 male) between 7 and 8 years (mean age = 7.8 years) old. The children were recruited at an elementary school in Kerkrade (Netherlands). The 3- to 4-year-old group originally consisted of 27 participants, but 7 children dropped out due to ocular or muscular artefacts or an insufficient number of trials in which they looked at the screen.

To ensure that all children performed in the normal cognitive range (IQ > 90) for their age, children completed an IQ test. All 3- to 4-year-old children successfully completed the SON–R 2½–7 (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998), and the 5- to 6-year-old and 7- to 8-year-old children performed two subtests of the WISC–III (Wechsler, 1991): the block design and vocabulary tests. The estimated total IQ score derived from these subtests has a mean reliability of .94 and a mean validity of .91 compared to the complete WISC–III (Spreen & Strauss, 1998). (See Table 1 for mean total IQ scores for each age group.) No children were excluded based on IQ. Additionally, to check for behavioral problems, parents completed the Child Behavior Checklist (CBCL; Achenbach, 1991). The CBCL is an instrument used for the detection of behavioral problems in children. In the 7- to 8-year-olds, 3 children (of 28) were excluded before experimental testing took place because they scored in the clinical range (> 63) on the Internalizing (I), Externalizing (E), and Total Problem subscales of the CBCL (Achenbach, 1991). Participants had no neurological history and had normal or corrected to normal vision as verified by their parents. In the Netherlands, all children get a standard visual checkup at the age of 3 or 4 years. All parents gave written informed consent for their child's participation in the study. A local ethical committee of the Faculty of Psychology at Maastricht University, Netherlands, approved the experimental procedure.

Table 1
Mean Full-Scale IQ Scores (SD in Parentheses) for the Three Age Groups

	Full scale IQ	Range
3–4 years	123.0 (9.3)	105–144
5–6 years	114.6 (10.2)	90–131
7–8 years	108.9 (9.9)	90–135

Note. Standard deviations are given in parentheses.

Stimuli and Task Procedure

Face stimuli consisted of 16 grayscale images (8 males, 8 females), half of which depicted a neutral expression, while the other half depicted a fearful expression. Fearful and neutral expressions were from different individuals. The photographs were taken from the NimStim Face Set (http://www.macbrain.org/resources.htm; Tottenham, Borscheid, Ellertsen, Marcus, & Nelson, 2002) and have been shown to evoke emotional effects at the level of the N170 in adults (Blau et al., 2007). Face images included European American and African American models. Fearful and neutral faces were equally balanced across race. Face pictures were trimmed to remove external features (neck and hairline). All pictures were fitted in a gray frame of 500 × 700 pixels. Each face subtended 6.3° of visual angle at a distance of 113 cm. The HSF images were created by filtering the original photographs using a high-pass cutoff of ≥ 6 cpd of visual angle (see Figure 1). The LSF images were created using a low-pass filter of ≤ 2 cpd of visual angle. Filtering was performed in Matlab (The Mathworks, Natick, MA) using a set of Gaussian filters.

The task consisted of four blocks, each containing 73 trials. Within each block, 64 faces and nine animation figures were presented on a gray background in randomized order. All faces were presented for 500 ms with an interstimulus interval of 1600–1800 ms. Participants were instructed to maintain fixation and to attend to all pictures. They had to press a response button as soon as they saw an animation figure on the screen and refrain from responding to all other images. This task was used to maintain the participant's attention during stimulus presentation. Short breaks were given between blocks. A video camera situated next to the screen recorded the child's gaze behavior. For 3- to 4-year-olds, looking behavior was coded offline, whereas it was monitored online for the 5- to 6-year-olds and 7- to 8-year-olds.

After the electroencephalogram (EEG) measurement, 5- to 6-year-old and 7- to 8-year-old children performed an additional reaction time task in which they had to decide as quickly and accurately as possible whether faces looked fearful or "usual" (neutral). Presentation time and interstimulus intervals were kept the same as in the initial ERP task. Participants started with a practice block that contained three phases. In the first phase, participants

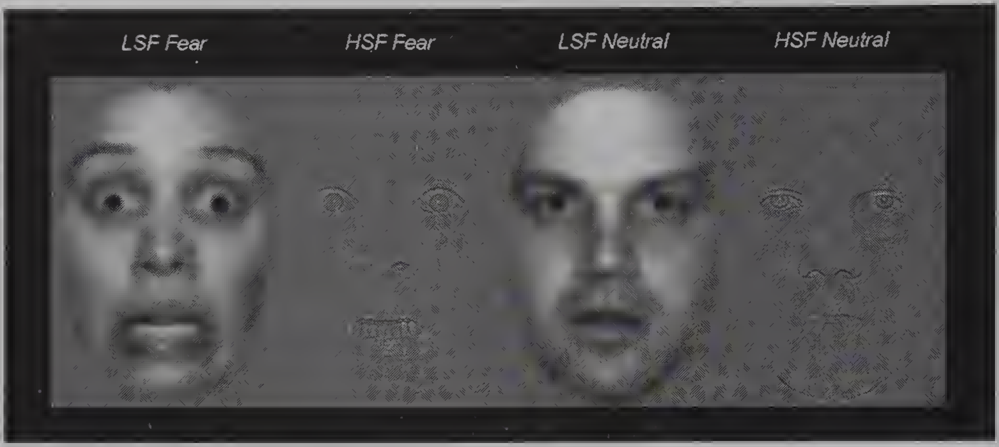


Figure 1. Examples of fearful and neutral low spatial frequency (LSF; < 2 cpd) and high spatial frequency (HSF; > 6 cpd) face stimuli.

were presented with a face of a lion with a fearful expression and were asked: ''How does the lion look?'' Incorrect answers were corrected by the experimenter. Thereafter participants were presented with a face of a lion that looked neutral and were asked: ''How does the lion look?'' Incorrect answers were again corrected by the experimenter. In the second phase, participants were presented with a block of 32 trials in which they had to decide whether the face of the lion looked ''usual'' or ''fearful'' by pressing a left- or right-hand button. The duration of stimulus presentation was locked to the button response. In the third phase, participants had to perform the same task as in Phase 2, but now with experimental task parameters: a fixed stimulus duration of 500 ms with an interstimulus interval of 1600–1900 ms. After the last phase, participants were told that they now had to perform a similar task that contained faces of humans instead of lion faces. Half of the stimuli (neutral) required a left button press whereas the other half (fearful) required a right button press. Left and right button presses were counterbalanced across participants. Right–left assignment was the same in the training and test conditions.

The task contained the same number of stimuli (256) as the initial task ERP but was divided into eight blocks. Furthermore, no animation figures were presented between trials.

ERP Recordings

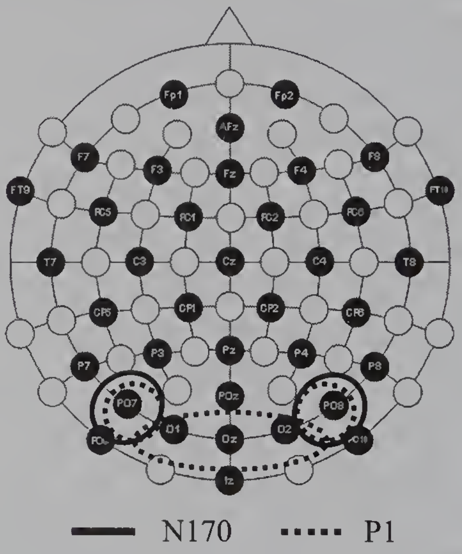


Figure 2. Scalp locations of the 36 electrodes.
Note. Mean amplitude and latency of the P1 were measured at pooled electrode sites: OZ/O1/O2/PO7/PO8. Mean amplitude and latency of the N170 were measured at occipito-temporal electrodes PO7/PO8.

Event-related brain potentials were recorded via a QuickCap (Compumedics, Victoria, Australia) containing 60 electrodes, of which 37 electrodes were used. The electrodes were placed according to the 10–20 system with intermediate positions (see Figure 2) and comprised Fp1/2, Fz, F3/4, F7/8, FC1/2, FC5/6, FT9/10, Cz, C3/4, T7/8, CP1/2, CP5/6, Pz, P3/4, P7/8, POz, PO7/8, Oz, O1/2, Iz, and PO9/10 plus a ground electrode at AFz.

Four additional electrodes placed above and below the left orbit and on the outer canthus of each eye were used to monitor vertical and horizontal eye movements. All electrodes were referenced to the left mastoid online and A2 was measured as an active electrode. Impedances were kept below 20 kΩ. A continuous EEG was recorded with a 500 Hz sampling rate and a band-pass filter of 0.01–200 Hz. The EEG data were analyzed offline using ''Vision Analyser'' software (Brain Products, Munich, Germany). The EEG data were epoched offline into 1500-ms periods, starting 200 ms prior to and ending 1300 ms after stimulus onset. Thereafter, the epochs were filtered (0.1–30 Hz), and artefacts from horizontal eye movements and blinks were reduced with the algorithm of Gratton, Coles,

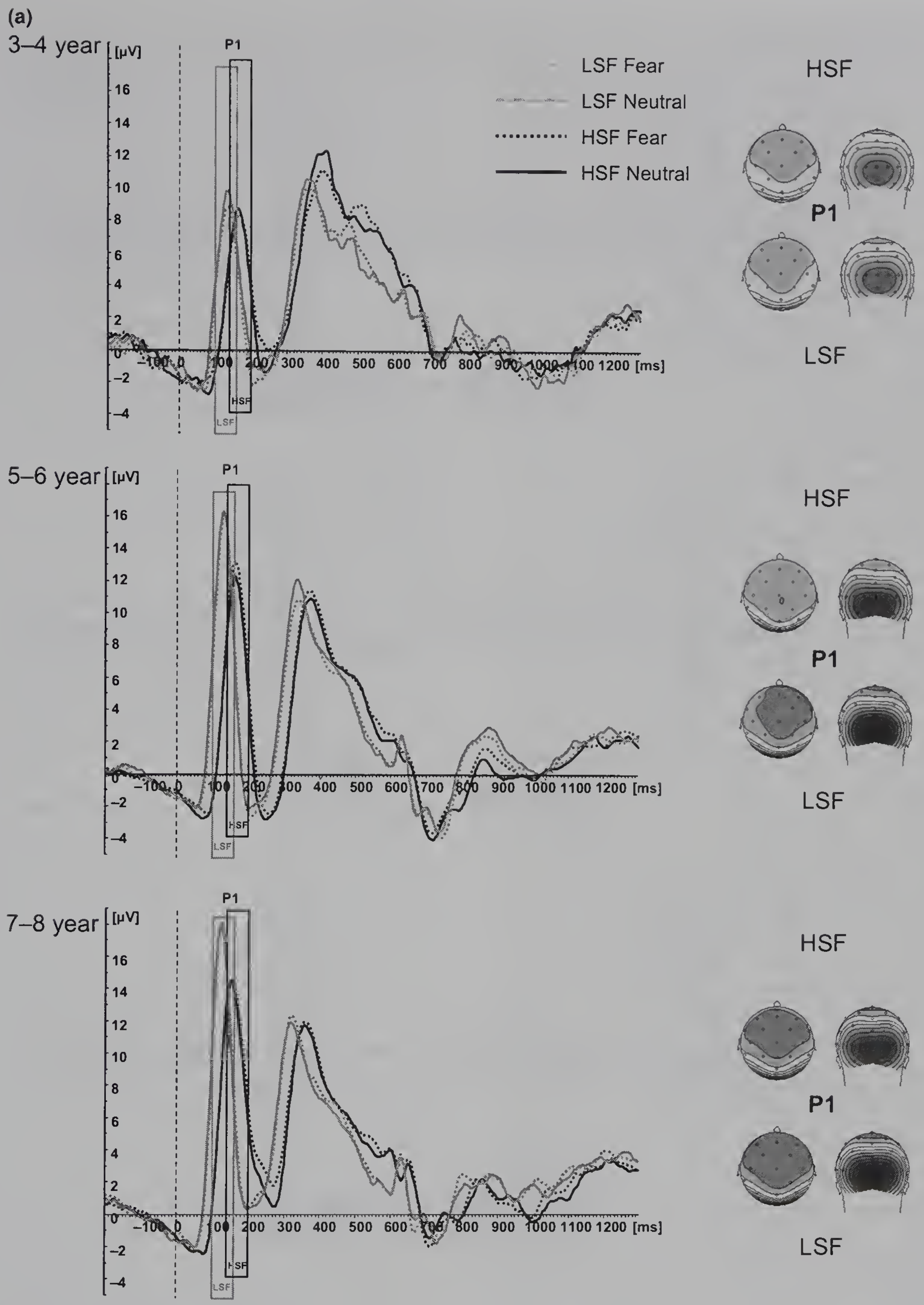


Figure 3. (a) Grand-average waveforms for 3- to 4-year-old, 5- to 6-year-old, and 7- to 8-year-old children at pooled electrodes (Oz/O1/O2/PO7/PO8), from which the P1 was extracted. In addition, the topographical distribution of the P1 is shown.

Note. The gray (low spatial frequency [LSF]) and black (high spatial frequency [HSF]) windows indicate the time windows over which P1 amplitude averages were calculated and P1 peak latency was extracted for statistical analysis. Note that because there is a difference between the way in which the grand-average waveforms (displayed in a and b) and the data used for statistical analysis are calculated (mean amplitudes across a predefined window; see Figure 4), the values in the current graphs may differ from the values used for statistical analysis (displayed in Figure 4).

(b) Grand-average waveforms for 3- to 4-year-old, 5- to 6-year-old, and 7- to 8-year-old children at electrodes PO7 and PO8, from which the N170 was extracted. In addition, topographical distributions of the N170 for HSF and LSF stimuli are shown.

Note. The gray (LSF) and black (HSF) windows indicate the time windows over which N170 amplitude averages were calculated and N170 peak latency was extracted for statistical analysis.

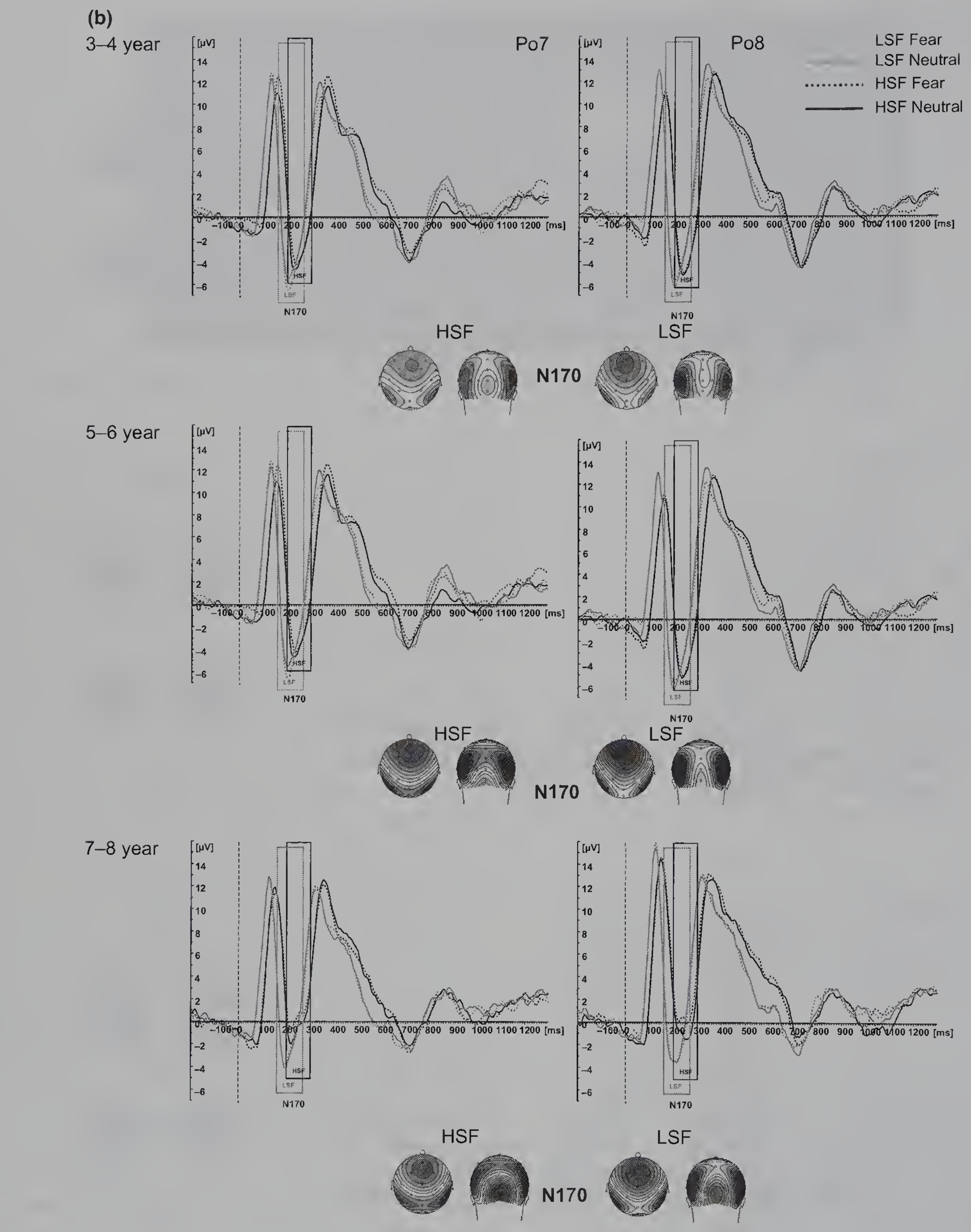


Figure 3. (Continued).

and Donchin (1983). After baseline correction (–200 to 0), trials that contained EEG artefacts (± 75 μV) were rejected from the data set. Furthermore, trials during which the child was not looking at the screen were discarded. Separate ERP averages were computed for all participants for the four stimulus conditions (SF: HSF, LSF) × (emotion: fear, neutral). Finally, the data were rereferenced to an average reference.

According to Cacioppo, Tassinary, and Berntson (2000), a minimum of 20 electrodes are sufficient for average references calculation.

Data Analysis

Based on topographies (spherical spline calculation) and peak maxima (see Figures 3a and 3b), regions or electrodes of interest were defined for

Table 2
Time Windows and Electrodes for Mean Amplitude and Latency Extraction for P1 and N170

ERP peak	LSF	HSF	Electrodes
P1	100–160	140–200	PO7, PO8, Oz, O1, O2
N170	170–270	200–300	PO7, PO8

Note. ERP = event-related brain potential; LSF = low spatial frequency; HSF = high spatial frequency.

both the P1 and N170 components (see Figure 2). Time windows for the components of interest (see Table 2) were chosen through visual inspection of the grand-average waveforms in all groups (see Figures 3a and 3b), which show a clear latency shift between LSF and HSF ERPs and show emotion effects (amplitude difference between fear and neutral) at the chosen time points. To measure P1 and N170 amplitudes, mean area amplitudes across a predefined time window (see Table 2) were used instead of peak amplitudes for two reasons. First, emotional effects were present across a broader time window and did not occur only at the peak (see Figures 3a and 3b). Second, mean areas are less sensitive to high-frequency noise then peak amplitudes (Luck, 2005).

The three groups differed in mean number of total trials included in the analyses, $F(2, 67) = 57.26$, $p < .001$, with fewer trials for the 3- to 4-year-olds (high neutral [HN] = 41, *SD* = 11, range = 25–64; high fear [HF] = 42, *SD* = 10, range = 27–62; low neutral [LN] = 43, *SD* = 9, range = 28–63; low fear [LF]: 43, *SD* = 10, range = 27–64) compared to the 5- to 6-year-olds (HN = 59, *SD* = 8, range = 34–64; HF = 59, *SD* = 7, range = 38–64; LN = 59, *SD* = 8, range = 33–64; LF = 59, *SD* = 8, range = 33–64) and 7- to 8-year-olds (HN = 61, *SD* = 5, range = 40–64; HF = 60, *SD* = 5, range = 43–64; LN = 60, *SD* = 5, range = 41–64; LF = 61, *SD* = 5, range = 48–64). Mean areas do not become biased when the noise level increases due to different trial numbers. Consequently, it is legitimate to compare mean amplitude measurements from waveforms based on different numbers of trials, whereas this is not legitimate for peak amplitude measurements (Luck, 2005). However, note that the lowest number of total trials included in the analysis ($n = 25$) can still be considered high. The mean amplitude of the P1 was extracted from a pooled set of occipito-temporal electrodes: PO7/PO8/Oz/O1/O2 (see Figure 2). The amplitude data were subjected to a 2 (SF: HSF, LSF) × 2 (emotion: fear, neutral) × 3 (group: 3- to 4-year-olds, 5- to 6-year-olds, 7- to 8-year-olds) analysis of variance (ANOVA). The mean amplitude of the N170 was extracted from electrodes PO7 and PO8 (see Figure 2). N170 data were subjected to a 2 (SF: HSF, LSF) × 2 (emotion: fear, neutral) × 2 (hemisphere: PO7, PO8) × 3 (group: 3- to 4-year-olds, 5- to 6-year-olds, 7- to 8-year-olds) ANOVA. Because the topographical distribution of the P1 did not show any laterality effects, we did not include this factor in the analysis. The location from which the different components were extracted is consistent with several ERP studies in children (see, e.g., Batty & Taylor, 2006; Dawson et al., 2004).

Because the grand averages (see Figures 3a and 3b) showed clear effects of SF on peak latency at the P1 and N170, we also analyzed peak latency for HSF and LSF. Although peak latencies, like peak amplitudes, are prone to high-frequency noise and have other shortcomings, there are not many good alternatives and so they are often the best measure (Luck, 2005). Based on the suggestions of Luck (2005), we took the following precautions: (a) We filtered out the high-frequency noise in the waveforms by filtering the preprocessed data with an additional 10-Hz low-pass filter (note, however, that using a 30-Hz low-pass filter, the latency results were replicated). (b) We used a peak detection algorithm that looked for local peak maxima rather than an absolute peak measure. When searching for local maxima, the automatic peak detection algorithm looks for extreme values within a given interval and considers the edge values only if this search fails. The difference between a local and an absolute maximum is that, in searching for an absolute maximum, the edge points of the intervals are found as peaks if the value there is greater (or less) than all values within the interval. (c) We only analyzed the effect of SF (which was clearly present in the data; see Figures 3a and 3b) in the different age groups, which resulted in a relatively high number of trials *per condition* (maximum = 128). Previous studies in adults have indicated that interactions between SF and emotion affect P1 and N170 amplitude only (Pourtois et al., 2005; Vlamings et al., 2009). P1 latency was extracted from a pooled set of occipito-temporal electrodes: PO7/PO8/Oz/O1/O2 (see Figure 2). P1 latency was subjected to a 2 (SF: HSF, LSF) × 3 (group: 3- to 4-year-olds, 5- to 6-year-olds, 7- to 8-year-olds) ANOVA. N170 latency was derived from electrodes PO7 and PO8 (see Figure 2) and these data were subjected to a 2 (SF: HSF, LSF) × 2 (hemisphere: PO7, PO8) × 3 (group: 3- to 4-year-olds, 5- to 6-year-olds, 7- to 8-year-olds) ANOVA.

Reaction times (only reaction times between 150 and 2100 ms after stimulus onset were included in the analysis) and percentage hits from the separate categorization task were subjected to a 2 (SF: HSF, LSF) × 2 (emotion: fear, neutral) × 2 (group: 5- to 6-year-olds, 7- to 8-year-olds) ANOVA. Four participants (three 5- to 6-year-olds and one 7- to 8-year-old) were excluded (significant outliers) from the analysis because they had too many misses.

When statistically significant first-, second-, or third-order interactions were found, these were further tested by looking at partial interactions in a fixed order (SF, emotion, hemisphere, and group). The higher order interactions had to be significant before lower order interactions were considered. When testing of significant interactions did not lead to significant results, these are not mentioned in the study.

Finally, for the 5- to 6-year-olds and 7- to 8-year-olds, correlation analyses were performed between behavioral measures (reaction times and percentage correct) and P1 and N170 amplitude and latency data in the different conditions.

Results

Grand averages and topographical distributions of the P1 and N170 are depicted in Figures 3a and 3b. The mean values and standard errors of mean amplitude and latency measures of the P1 and N170 used for statistical analysis are depicted in Figures 4 and 5.

P1 Amplitude

The overall four-way ANOVA revealed an interaction between SF and emotion, $F(1, 67) = 5.42$, $p < .05$, $\eta^2_p = .08$. Pairwise comparisons indicated significantly higher P1 amplitudes for fearful compared to neutral faces presented at HSF, $t(69) = 2.61$, $p < .05$, $\eta^2_p = .09$ (see Figures 3a and 4), whereas the difference between neutral and fearful faces was not significant for LSF faces. Furthermore a main effect of group was found irrespective of SF, $F(2, 67) = 10.22$, $p < .001$, $\eta^2_p = .23$, showing that P1 amplitudes of 3- to 4-year-old children were larger than those in 5- to 6-year-old children, $t(43) = -3.65$, $p < .01$, Cohen's $d = -1.39$, and 7- to 8-year-old, $t(43) = -4.73$, $p < .001$, Cohen's $d = -1.14$ (see Figures 3a and 5). Other main effects were not found.

P1 Latency

The analysis of peak latency of P1 only revealed a significant main effect of SF, $F(1, 67) = 549.59$, $p < .001$, $\eta^2_p = .89$, with HSF faces showing longer P1 latencies than LSF faces (see Figures 3a and 5).

N170 Amplitude

The overall four-way ANOVA revealed a significant interaction between SF and emotion, $F(1, 67) = 9.20$, $p < .01$, $\eta^2_p = .10$. There were no main effects. As for P1, pairwise comparisons indicated a significant effect of emotion for HSF faces only. For the N170, however, neutral HSF faces elicited significantly higher amplitudes than HSF fearful faces, $t(69) = -3.67$, $p < .001$, $\eta^2_p = .11$, whereas there was no effect of emotion for LSF faces (see Figures 3b and 4).

N170 Latency

For N170 peak latency, we only found a significant SF × Group interaction, $F(2, 67) = 4.42$, $p < .05$, $\eta^2_p = .12$ and no main effects. The SF effect was significant in each of the age groups, with LSF faces showing shorter peak latencies than HSF faces: 3- to 4-year-olds, $t(19) = -5.04$, $p < .001$, $\eta^2_p = .57$; 5- to 6-year-olds, $t(24) = -13.09$, $p < .001$, $\eta^2_p = .88$; 7- to 8-year-olds, $t(24) = -18.34$, $p < .001$, $\eta^2_p = .93$ (see Figures 3b and 5). Further analysis revealed a significant effect of group for LSF faces only, $F(2, 67) = 3.68$, $p < .05$. Three- and 4-year-old children had longer N170 latencies than 7- to 8-year-old children, $t(43) = 2.54$, $p < .05$, $\eta^2_p = .10$ (see Figures 3b and 5), when presented with an LSF face, whereas there was no difference between the 3- to 4-year-olds and 5- to 6-year-olds. Post hoc correlation analyses indicated a significant negative correlation between age in months (including all age groups) and latency for LSF faces $r = -.27$, $p < .05$. This effect was absent for HSF faces.

Percentage Hits—Categorization Task

Analysis of the percentage hits recorded in the categorization task in the 5- to 6-year-olds and 7- to 8-year-olds indicated a marginally significant Group × Emotion interaction, $F(1, 44) = 3.99$, $p = .052$, $\eta^2_p = .08$; 7- to 8-year-olds performed slightly better on neutral compared to fearful faces, $t(23) = -3.02$, $p < .01$, $\eta^2_p = .28$, but there was no stimulus difference for 5- to 6-year-olds. When testing group differences for neutral and fearful stimuli, 5- to 6-year-olds had fewer hits in the neutral condition compared to 7- to 8-year-olds, $t(44) = -2.58$, $p < .05$, $\eta^2_p = .13$. There was a significant positive correlation between age in months and percentage

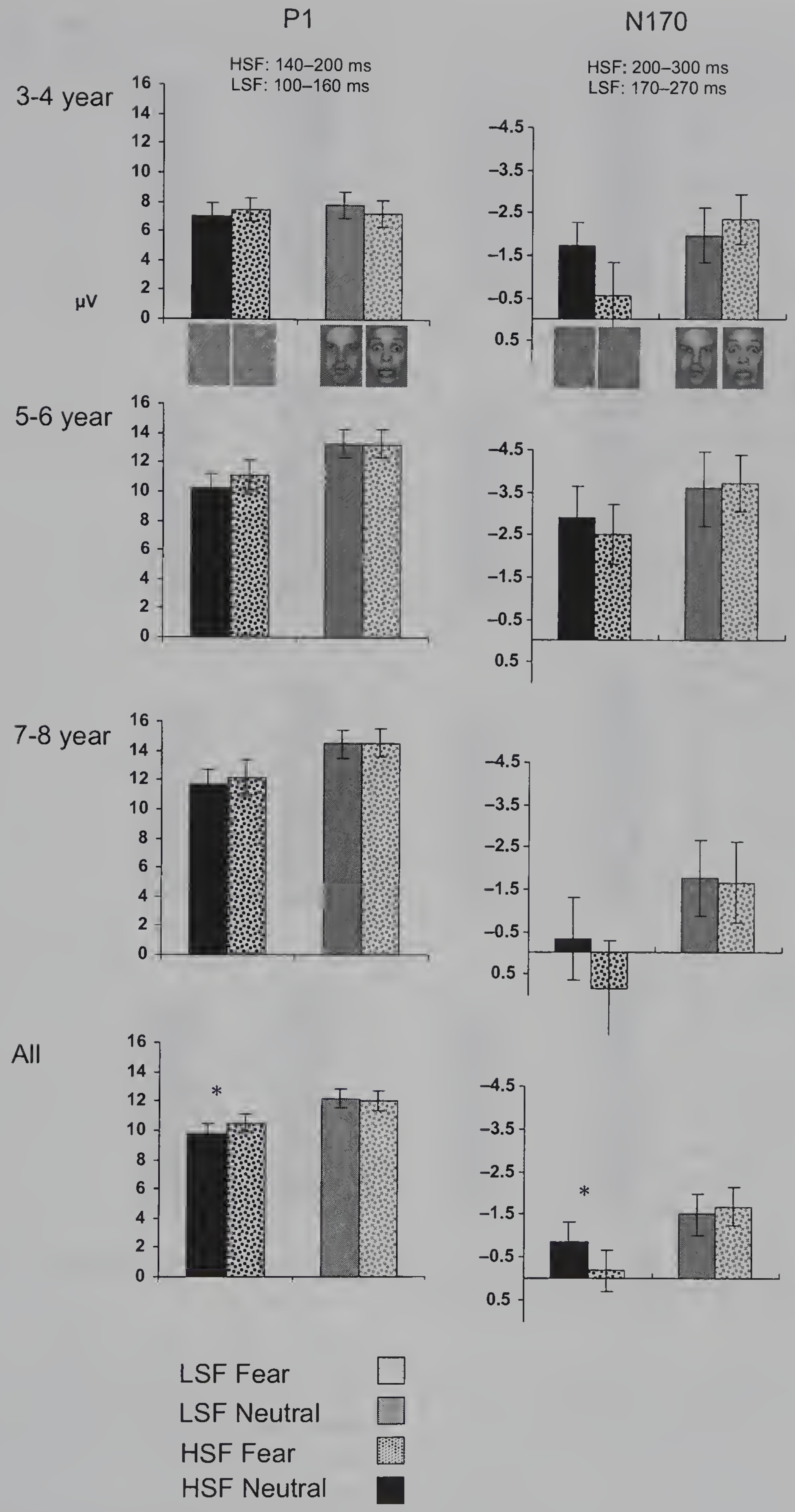


Figure 4. Bar graphs of the mean amplitudes (+*SE*) of the P1 and N170 in all stimulus conditions (high spatial frequency [HSF] fear, HSF neutral, low spatial frequency [LSF] fear, LSF neutral) are shown separately for each age group.

Note. Note that because there is a difference in the way in which the grand averages (displayed in Figures 3a and 3b) and the data used for statistical analysis, shown in the bar graphs in this figure, are calculated, the values in the current graphs may differ from the values displayed in Figures 3a and 3b. In the top row, the windows over which the amplitude averages were calculated are reported for P1 and N170.

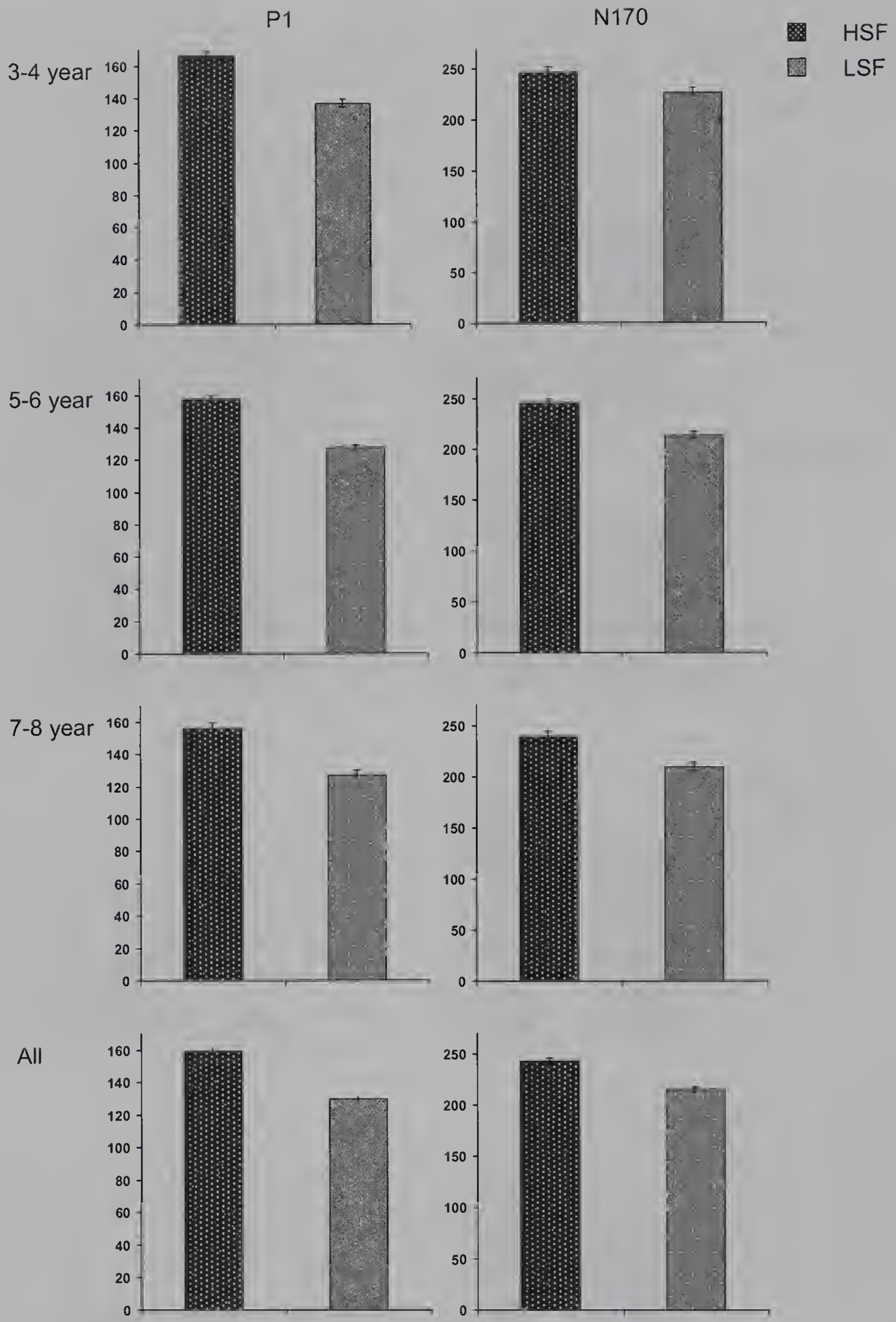


Figure 5. Bar graphs of the mean latencies (+*SE*) of the P1 and N170 in all stimulus conditions (high spatial frequency [HSF] and low spatial frequency [LSF]) are shown separately for each age group.

of correct trials in the neutral condition, $r = .35$, $p < .05$, but not in the fearful condition.

Reaction Times—Categorization Task

Analysis of the reaction times revealed a marginally significant effect of emotion, $F(1, 44) = 3.96$, $p = .053$, $\eta^2_p = .08$, indicating faster reaction times to recognize fearful than neutral faces (see Table 3). In addition, we found a main effect of SF, $F(1, 44) = 39.69$, $p < .001$, $\eta^2_p = .74$. Participants were faster, overall, in categorizing LSF faces than HSF faces (see Table 3).

Correlation Between Behavioral Performance and ERP Measures

In none of the conditions were significant correlations between reaction times or percentage hits and P1/N170 latency or amplitude found.

Table 3
Mean Reaction Times (RT) and Percentage Hits for All Stimulus Conditions Averaged Across 5- to 6-Year-Old and 7- to 8-Year-Old Children

	LSF		HSF	
	Neutral	Fear	Neutral	Fear
Hit RT (years)				
5–6	1001 (45)	988 (44)	1030 (43)	1019 (44)
7–8	986 (29)	961 (33)	1026 (31)	1002 (31)
Hit percentage (years)				
5–6	70 (5)	75 (3)	69 (4)	72 (4)
7–8	83 (3)	81 (3)	81 (3)	75 (3)

Note. Standard errors are given in parentheses. LSF = low spatial frequency; HSF = high spatial frequency.

Discussion

The aim of this study was to investigate the effect of manipulation of the spatial frequency content (LSF and HSF) of face stimuli on the rapid processing of negative (fearful) and neutral facial expressions in the visual cortex of 3- to 8-year-old children. Recent studies have revealed that, in adults, early effects of facial emotion content (< 200 ms) on ERP components above the visual cortex (P1 and N170), are primarily mediated by LSF information (Pourtois et al., 2005; Vlamings et al., 2009).

In contrast, in this study we found that the early effects of emotion expression detection at the P1 were driven primarily by more detailed HSF information in all children (such as contours of the eyes, eyebrows, mouth, and so on). This was shown by higher P1 amplitudes for fearful compared to neutral faces for HSF images only. Similar enhanced P1 amplitudes in response to broadband fearful face stimuli (containing all frequencies) are typically interpreted as reflecting involuntary capture of attention by fear (e.g., Pourtois et al., 2004, 2005). Whereas this Emotion × SF interaction effect was found for all children, a group main effect indicated that P1 amplitude decreased with age, consistent with Batty and Taylor (2003, 2006). These authors suggested that this developmental effect reflects automatization of visual processing in general. However, main effects of group on ERP amplitude are difficult to interpret and might also reflect structural brain changes. To exclude the possibility that the P1 findings were influenced by unintended differences in low-level visual characteristics between neutral and fearful faces resulting from filtering, additional analyses were performed. These analyses indicated that the spatial frequency content of the two expression types did not differ substantially, making it highly unlikely that such differences influenced the present results. Furthermore, the present results show that besides for P1, also for the face-specific N170, effects of emotional expression are driven by a different type of SF information compared to adults (Vlamings et al., 2009).

A possible explanation for this developmental effect must take into account findings that consistently show that infants just a few months old already have a "preference" for fearful faces, as shown by longer looking times (Kotsoni, de Haan, & Johnson, 2001; Nelson & Dolgin, 1985; Peltola, Leppänen, Mäki, & Hietanen, 2009) and larger ERP responses to fearful than other faces (de Haan, Belsky, Reid, Volein, & Johnson, 2004; Leppänen, Moulson, Vogel-Farley, & Nelson, 2007; Nelson & de Haan, 1996). This early preference for fear is most probably mediated by LSF cues since, due to immaturity of the visual system, 6- to 12-month-old infants have very little sensitivity to HSF (de Heering et al., 2008; Slater & Sykes, 1977). Integrating this literature with the present data yields the following developmental pattern: During infancy, the early visual processing of fearful expressions is (most strongly) mediated by LSF cues, followed by a shift to HSF cues in childhood as is shown by the present ERP data, then returning to the use of LSF cues in adulthood (Pourtois et al., 2005; Vlamings et al., 2009).

The enhanced use of HSF cues for early visual processing of fear during childhood, as opposed to LSF cues in adults, might be linked to immaturity of the subcortical face-processing route (see the Introduction). This route makes use of LSF and is suggested to boost P1 amplitudes to fearful LSF faces through direct modulating feedback signals coming from the amygdala (Pourtois et al., 2005). In support of this hypothesis, Cunningham, Bhattacharyya, and Benes (2002) found that amygdala–cortical connectivity continues to mature into adolescence, at least in the prefrontal areas. However, such an explanation would be difficult to reconcile with findings suggesting that the subcortical face-processing route is already functional in newborns and probably plays a crucial role in the bias for fearful expressions that is already present at this age (see for review Johnson, 2005; Morton & Johnson, 1991). Finally, the conclusion that early LSF-based processing of fearful expressions in adults is mediated by a rapid subcortical processing route is, as yet, only derived from indirect evidence (see Pourtois et al., 2005; Vuilleumier et al., 2003). It is possible that this early, LSF-based, facial emotion processing in the visual cortex results from a rapid feedforward cortical projection within the visual system itself. Future studies, possibly using fMRI (or a combination of fMRI and EEG), should be aimed at further disentangling these possibilities.

An alternative explanation that takes into account both the maturation of the visual system and the influence of developmental differences in expertise with face processing might better fit the data. As the ability to resolve HSF only reaches maturity at 3–4 years (e.g., Adams & Courage, 2002), children presumably might not be able to process relevant HSF emotional cues before that age. This is supported by findings showing that infants use mainly LSF cues for face processing (de Heering et al., 2008). At the age of 3–4 years, however, the ability to resolve HSF information reaches maturity, and HSF information might become most efficient for the rapid extraction of emotion, as found in this study at least up to 8 years. The developmental shift to the use of LSF cues again in adulthood, as seen in both P1 and N170, might in turn be related to increased expertise with facial expressions and faces in general.

Some evidence for such an expertise account comes from a study by Schyns and Oliva (1999), showing that adults more frequently used LSF information for face identification when they were familiarized with faces by training. A similar effect was found for object identification by Viggiano, Righi, and Galli (2006). In addition, several studies show stronger configural processing (i.e., a larger inversion effect) for face stimuli for which participants have expertise (see for review McCleery et al., 2008; Parr & Heintz, 2008), and LSFs are known to play an important role in such configural face processing (Flevaris, Robertson, & Bentin, 2008; Goffaux, Gauthier, & Rossion, 2003; Goffaux, Hault, Michel, Vuong, & Rossion, 2005).

Surprisingly, the HSF emotion effect in the 3–8-year-old children at the N170 was opposite to that of the P1. The P1 data suggest increased processing of fearful HSF stimuli, while, at the N170, amplitudes were larger in response to neutral than fearful HSF stimuli. This might be due to ambiguity of neutral faces for young children. A recent study showed that before the age of 9 years, children often rate neutral faces as happy or sad (Durand, Gallay, Seigneuric, Robichon, & Baudouin, 2007; see also Carlson, Felleman, & Masters, 1983; Felleman, Barden, Carlson, Rosenberg, & Masters, 1983; Reichenbach & Masters, 1983). In this study, the oldest (7–8 years old) children also correctly categorized a higher percentage of neutral faces than 5- to 6-year-olds. Additionally, Thomas et al. (2001) demonstrated greater amygdala activity in response to neutral faces than fearful faces in children than adults (see also Tottenham, Hare, & Casey, 2009). It was argued that this enhanced activity might signal increased effort invested in the decoding of the neutral face expression that is perceived as ambiguous by children. Such an explanation would be congruent with the fact that amplitudes to neutral faces were only enhanced for the N170 and not the P1 in this study. As opposed to the P1, which is assumed to be primarily sensitive to low-level visual processing or attentional manipulations, the N170 is thought to be especially involved in structural encoding of faces (Jacques & Rossion, 2006; Nobre, Rao, & Chelazzi, 2006; Taylor, 2002; Valdes-Sosa, Bobes, Rodriguez, & Pinilla, 1998).

With respect to the latencies of the P1 and N170, we found faster latencies for processing of LSF compared to HSF stimuli, irrespective of facial expression, which is consistent with findings in adults for facial (Halit, de Haan, Schyns, & Johnson, 2006; Hsiao, Hsieh, Lin, & Chang, 2005; McCarthy, Puce, Belger, & Allison, 1999) and nonfacial (Mihaylova, Stomonyakov, & Vassilev, 1999; Musselwhite & Jeffreys, 1985) stimuli. Similar effects were found with respect to reaction times in the current behavioral categorization task. The temporal precedence of LSF is consistent with findings that the neuronal pathways sensitive to LSF and HSF have dissociable time scales, with faster cortical arrival of information that is processed in the LSF-sensitive magnocellular pathway (see for review Laycock, Crewther, & Crewther, 2007).

Importantly, the latency of the N170 decreased in the LSF condition with increasing age, whereas there were no significant age effects for processing of HSF faces. Faster N170 latencies with increasing age have previously been reported for face stimuli containing all SF frequencies (see for review Taylor, Batty, & Itier, 2004). This study thus suggests that this decrease in N170 latency to unfiltered faces might be caused specifically by faster extraction of LSF information. In contrast to these developmental effects on N170 latency for LSF, no age effects were found with respect to P1 and N170 amplitude. That is, early effects of emotion at P1 and N170 amplitude were primarily driven by HSF, and this did not change with age. Integration of the latency and amplitude results suggests that although children seem to more rapidly extract LSF information for face processing with age, this information does not yet drive emotional processing during early face-processing stages.

Finally, in a behavioral task, we investigated whether an LSF or HSF advantage for the processing of facial expressions would be reflected in reaction times to an active categorization task. Like adults (Vlamings et al., 2009), children decided more quickly that a face was fearful than that it was neutral, which is consistent with the finding that stimuli that signal threat receive preferential attention over neutral stimuli (see for review Holmes, Green, et al., 2005). The lack of an interaction between emotion and SF suggests that emotional expressions are recognized based on LSF as well as HSF information at later information processing stages (at response level). This is supported by data from another recent behavioral study that indicated that both HSF and LSF cues are used for emotion identification in typically developing children (one age group: 6–16 years; Deruelle, Rondan, Salle-Collemiche, Bastard-Rosset, & Da Fonséca, 2007). We did not find evidence for an association between behavioral performance and ERP measures suggesting that in this study early and late stages of processing might not be directly linked.

In conclusion, the present data show that just as in adults, 3- to 8-year-old children show enhanced early visual brain responses (P1) to fearful compared to neutral faces. However, whereas prior work showed that this effect was mediated by LSF cues in adults and infants, present data showed that in children more detailed (HSF) face cues were used (such as eyes, mouth, and so on). The absence of an LSF bias for the processing of fear in 3- to 8-year-olds might be explained by a combination of immaturity of the visual system to process HSF information and a lack of expertise with face processing that enables adults to rely on LSF information. Future studies with older age groups should establish when the processing of fear starts to be driven primarily by LSF information. Finally, our data suggest that a developmental latency decrease of the N170 could be linked to faster extraction of LSF cues with age. The present data might be important for research in neurodevelopmental disorders such as autism, for which specific deficits in the processing of LSF have been suggested (see for review Johnson, 2005; Laycock et al., 2007) and might lead to an investigation of how this LSF bias in autism is linked to the development of gaze behavior and face expertise. Finally, it should be noted that due to practical considerations, the present study only included one emotional expression (fear) and that the present effects thus cannot be generalized to general (negative) emotion processing, which should be the subject of a future study.

References

Acerra, F., Burnod, Y., & de Schonen, S. (2002). Modelling aspects of face processing in early infancy. *Developmental Science*, *5*, 98–117.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4–18*. Burlington: Department of Psychiatry, University of Vermont.

Adams, J. A., & Courage, M. L. (2002). Using a single test to measure human contrast sensitivity from early childhood to maturity. *Vision Research*, *42*, 1205–1210.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*, 613–620.

Batty, M., & Taylor, M. J. (2006). The development of emotional face processing during childhood. *Developmental Science*, *9*, 207–220.

Blau, V. C., Maurer, U., Tottenham, N., & McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, *3*, 7.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology*. Cambridge, UK: Cambridge University Press.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience*, *15*, 210–227.

Carlson, C. R., Felleman, E. S., & Masters, J. C. (1983). Influence of children's emotional states on the recognition of emotion in peers and social motives to change another's emotional state. *Motivation and Emotion*, *7*, 61–79.

Cunningham, M. G., Bhattacharyya, S., & Benes, F. M. (2002). Amygdalo-cortical sprouting continues into early adulthood: Implications for the development of normal and abnormal function during adolescence. *Journal of Comparative Neurology*, *453*, 115–130.

Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science*, *7*, 340–359.

de Haan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal of Child Psychology and Psychiatry*, *45*, 1209–1218.

de Haan, M., Nelson, C. A., Gunnar, M. R., & Tout, K. A. (1998). Hemispheric differences in brain activity related to the recognition of emotional expressions by 5-year-old children. *Developmental Neuropsychology*, *14*, 495–518.

de Heering, A., Turati, C., Rossion, B., Bulf, H., Goffaux, V., & Simion, F. (2008). Newborns' face recognition is based on spatial frequencies below 0.5 cycles per degree. *Cognition*, *106*, 444–454.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

Deruelle, C., & Fagot, J. (2005). Categorizing facial identities, emotions, and genders: Attention to high- and low-spatial frequencies by children and adults. *Journal of Experimental Psychology*, *90*, 172–184.

Deruelle, C., Rondan, C., Salle-Collemiche, X., Bastard-Rosset, D., & Da Fonséca, D. (2007). Attention to low- and high spatial frrequencies in categorizing facial identities, emotions and gender in children with autism. *Brain and Cognition*, *66*, 115–123.

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, *97*, 14–27.

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport*, *25*, 427–431.

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research*, *39*, 2325–2333.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, *187*, 517–552.

Felleman, E., Barden, R. C., Carlson, C. R., Rosenberg, L., & Masters, J. C. (1983). Children's and adults' recognition of spontaneous and posed emotional expressions in young children. *Developmenal Psychology*, *19*, 405–413.

Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing of face features and face configuration: An ERP analysis. *Brain Research*, *1194*, 100–109.

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Brain Research*, *16*, 416–424.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies is supporting configural and featural processing of faces. *Perception*, *34*, 77–86.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1023–1039.

Goldstein, E. B. (1999). *Sensation and perception*. Pacific Grove, CA: Brooks/Cole.

Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, *46*, 1253–1262.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for offline removal of ocular artifact. *Electroencephalogram Clinical Neurophysiology*, *55*, 468–484.

Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*, 154–161.

Hammarrenger, B., Leporél, F., Lippé, S., Labrosse, M., Guillemot, J. P., & Roy, M. S. (2003). Magnocellular and parvocellular developmental course in infants during the first year of life. *Documenta Ophthalmologica*, *107*, 225–233.

Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., & Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral Cortex*, *13*, 793–805.

Hess, R. F. (2004). Spatial scale in visual processing. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1043–1059). Cambridge, MA: MIT Press.

Holmes, A., Green, S., & Vuilleumier, P. (2005). The involvement of distinct visual channels in rapid attention towards fearful facial expressions. *Cognition and Emotion*, *19*, 899–922.

Holmes, A., Winston, J. S., & Eimer, M. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Cognitive Brain Research*, *25*, 508–520.

Hsiao, F. J., Hsieh, J. C., Lin, Y. Y., & Chang, Y. (2005). The effects of face spatial frequencies on cortical processing revealed by magnetoencephalography. *Neuroscience Letters*, *380*, 54–59.

Jacques, C., & Rossion, B. (2004). Concurrent processing reveals competition between visual representations of faces. *Neuroreport*, *15*, 2417–2421.

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychological Science*, *17*, 485–492.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*, 766–774.

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception*, *30*, 1115–1125.

Laycock, R., Crewther, S. G., & Crewther, D. P. (2007). A role for the magnocellular advantage in visual impairments in neurodevelopmental and psychiatric disorders. *Neuroscience and Biobehavioral Reviews*, *31*, 363–376.

Le Doux, J. E. (1996). *The emotional brain*. New York: Simon & Schuster.

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development*, *78*, 232–245.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

McCarthy, G., Puce, A., Belger, A., & Allision, T. (1999). Electrophysiological studies of human face perception II: Response properties of face specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, *9*, 431–444.

McCleery, J. P., Zhang, L., Ge, L., Wang, Z., Christiansen, E. M., Lee, K., et al. (2008). The roles of visual expertise and visual input in the face inversion effect: Behavioral and neurocomputational evidence. *Vision Research*, *48*, 703–715.

Mihaylova, M., Stomonyakov, V., & Vassilev, A. (1999). Peripheral and central delay in processing high spatial frequencies: Reaction time and VEP latency studies. *Vision Research*, *39*, 699–705.

Morison, V., & Slater, A. (1985). Contrast and spatial frequency components in visual preferences of newborns. *Perception*, *14*, 345–348.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8*, 454–469.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review, 98*, 164–181.

Musselwhite, M. J., & Jeffreys, D. A. (1985). The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. *Vision Research, 25*, 1545–1555.

Nelson, C. A., & de Haan, M. (1996). Neural correlates of infants'' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology, 29*(7), 577–595.

Nelson, C. A., & Dolgin, K. (1985). The generalized discrimination of facial expressions by 7-month-old infants. *Child Development, 56*, 58–61.

Nobre, A. C., Rao, A., & Chelazzi, L. (2006). Selective attention to specific features within objects: Behavioral and electrophysiological evidence. *Journal of Cognitive Neurocience, 18*, 539–561.

Parr, L. A., & Heintz, M. (2008). Discrimination of faces and houses by rhesus monkeys: The role of stimulus expertise and rotation angle. *Animal Cognition, 11*, 467–474.

Peltola, M. J., Leppänen, J. M., Mäki, S. M., & Hietanen, J. K. (2009). *E*mergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience, 4*, 134–142.

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping. *Human Brain Mapping, 26*, 65–79.

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex, 14*, 619–633.

Reichenbach, L., & Masters, J. C. (1983). Children's use of expressive and contextual cues in judgements of emotion. *Child Development, 54*, 993–1004.

Rossion, B., Campanella, S., Gomez, C. M., Delinte, A., Debatisse, D., Liard, L., et al. (1999). Task modulation of brain activity related to familiar and unfamiliar face processing: An ERP study. *Clinical Neurophysiology, 110*, 449–463.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69*, 243–265.

Skottun, B. C., & Skoyles, J. R. (2007). A few remarks on relating reaction time to magnocellular activity. *Journal of Clinical and Experimental Neuropsychology, 29*, 860–866.

Slater, A., & Sykes, M. (1977). Newborn infants' visual responses to square wave gratings. *Child Development, 48*, 545–554.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms and commentary*. New York: Oxford University Press.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. *Neuroreport, 15*, 777–780.

Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clinical Neurophysiology, 113*, 1903–1908.

Taylor, M. J., Batty, M., & ltier, R. J. (2004). The faces of development: A review of early face processing over childhood. *Journal of Cognitive Neuroscience, 16*, 1426–1442.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Manual and justification of the SON-R 2.5–7*. Göttingen, Germany: Hogrefe-Verlag.

Thomas, K. M., Drevets, W. C., Whalen, P. J., Eccard, C. H., Dahl., R. E., Ryan, N. D., et al. (2001). Amygdala response to facial expressions in children and adults. *Biological Psychiatry, 49*, 309–316.

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D. J., & Nelson, C. A. (2002). Categorization of facial expressions in children and adults: Establishing a larger stimulus set. *Journal of Cognitive Neuroscience, S74*.

Tottenham, N., Hare, T. A., & Casey, B. J. (2009). A developmental perspective on human amygdala function. In E. Phelps & P. Whalen (Eds.), *The human amygdala*. (pp. 107–118). New York: Guilford.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience, 10*, 137–151.

Viggiano, M. P., Righi, S., & Galli, G. (2006). Category-specific visual recognition as affected by aging and expertise. *Archives of Gerontology and Geriatrics, 42*, 329–338.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? *Journal of Vision, 9*, 1–13.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6*, 624–631.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition. Manual*. San Antonio: Psychological Corporation.

Williams, L. M., Liddell, B. J., Rathjen, J., Brown, K. J., Gray, J., Phillips, M., et al. (2004). Mapping the time course of nonconscious and conscious perception of fear: An integration of central and peripheral measures. *Human Brain Mapping, 21*, 64–74.

Winston, J. S., Vuilleumier, P., & Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Current Biology, 13*, 1824–1829.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1320–1335

Typologies of Family Functioning and Children's Adjustment During the Early School Years

Melissa L. Sturge-Apple
and Patrick T. Davies
University of Rochester

E. Mark Cummings
University of Notre Dame

Guided by family systems theory, the present study sought to identify patterns of family functioning from observational assessments of interparental, parent–child, and triadic contexts. In addition, it charted the implications for patterns of family functioning for children's developmental trajectories of adjustment in the school context across the early school years. Two-hundred thirty-four kindergarten children (129 girls and 105 boys; mean age = 6.0 years, *SD* = 0.50 at Wave 1) and their parents participated in this multimethod, 3-year longitudinal investigation. As expected, latent class analyses extracted 3 primary typologies of functioning including: (a) cohesive, (b) enmeshed, and (c) disengaged families. Furthermore, family patterns were differentially associated with children's maladaptive adjustment trajectories in the school context. The findings highlight the developmental utility of incorporating pattern-based approaches to family functioning.

Homes characterized by high levels of interparental discord evidence greater perturbations in parenting and emotional unavailability (e.g., Sturge-Apple, Davies, & Cummings, 2006), as well as overreactive discipline and difficulties controlling or managing children (Lorber & O'Leary, 2005; Stoneman, Brody, & Burke, 1989). In addition, children in families with high levels of interparental discord and parenting difficulties have greater problems in meeting developmental milestones (e.g., Gonzales, Pitts, Hill, & Roosa, 2000; Sturge-Apple et al., 2006; Webster-Stratton & Hammond, 1999). Much of the work from this field of inquiry has utilized variable-based approaches in delineating associations between interparental and parent–child relationships and children's development. However, the moderate relations between interparental and parent–child functioning highlight that there is not a one-to-one correspondence in difficulties across family subsystems (e.g., Erel & Burman, 1995). Thus, although variable-based models of family functioning have been informative, exclusive reliance on linear continua to characterize functioning across family subsystems does not fully capture qualitatively different profiles of communication patterns across family subsystems (Minuchin, 1985). To overcome these limitations, researchers have increasingly called for the utilization of pattern-based approaches toward also identifying nonlinear configurations of relationship qualities across family subsystems (e.g., Belsky & Fearon, 2004; Johnson, 2003; McHale, Kazali, Rotman, Talbot, Carleton, and Lieberson, 2004; O'Connor, Hetherington, & Reiss, 1998). Pattern-based approaches, such as those used in cluster-analytic and latent class techniques, reveal qualitatively different profiles of study variables that are not anchored on a linear or continuous scale.

Family systems theory challenges researchers to extend beyond understanding child development solely in the context of the parent–child relationship by incorporating collective experiences in the broader family unit in models of family process. However, applying family systems theory to developmental research designs remains a formidable challenge (McHale, Kuersten-Hogan, & Rao, 2004). At a conceptual level, employing the family systems paradigm in understanding children's

---

This research was supported by a grant from the National Institute of Mental Health (R01 MH57318) awarded to Patrick T. Davies and E. Mark Cummings. Melissa Sturge-Apple was supported by the National Institute of Nursing Research (R21 NR010857). The authors are grateful to the children, parents, teachers, and school administrators who participated in this project. Their gratitude is also expressed to project staff, including: Courtney Forbes, Courtney Henry, Marcie Goeke-Morey, Amy Keller, Michelle Sutton, and the graduate and undergraduate students at the University of Rochester and University of Notre Dame.

Correspondence concerning this article should be addressed to Melissa L. Sturge-Apple, Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY 14627. Electronic mail may be sent to mapple@psych.rochester.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8104-0020

developmental difficulties necessitates accounts of child adaptation within a rich, but parsimonious, characterization of the interplay among relationships and individuals in the family context (e.g., Cowan & Cowan, 2002; McHale, Kuersten-Hogan, & Lauretti, 1996). By the same token, the incorporation of methodological and analytic advances in studies examining developmental models of family systems has lagged behind as researchers have tended to utilize variable-based methods rather than pattern-based methods to examine interplay among family subsystems. Accordingly, the aims of the present study are to: (a) utilize a pattern-based approach to identify family typologies across multiple family contexts and (b) examine how family typologies are associated with children's socioemotional adjustment in extrafamilial settings, especially the school context.

A central assumption of family systems theory is that interdependencies among relationships within the family are governed by boundaries or implicit rules for accessing materials, resources, and support within the family. Guided by the concept of boundaries, family systems theorists have consistently identified three qualitatively distinct profiles of family interactions characterized by harmony, disengagement, and enmeshment, respectively (e.g., Minuchin, 1974). Harmonious or cohesive families have well-defined, yet permeable boundaries that permit children access to resources (e.g., support, warmth) without undermining relational autonomy with other family members (Cox & Paley, 1997). Relatively warm, emotionally close and agreeable relationships across multiple family subsystems characterize this family type, with negative affect and difficulties effectively managed to prevent spillover from any single subsystem into other family relationships. In contrast, disengaged families have rigid boundaries, manifested in cold, indifferent, unsupportive, and emotionally withdrawn family relationships. Communication across family subsystems is stymied and difficult and family members function as distinct entities rather than part of a unified whole. Finally, the diffuse, thin boundaries of enmeshed families are reflected in emotionally seamless and entangled relationships. Children may experience some degree of warmth and support in these family relationships, but access to these resources occurs at a cost, including the proliferation of hostility and distress from one family subsystem to another, intrusive relationships, and significant restrictions in personal and psychological autonomy. Although other possible family systems are sometimes described in this literature, our tri-partite classification scheme is consistent with Minuchin's conceptualizations stressing boundary determination and differentiation of separate family subsystems (Minuchin, 1974), supported by both clinical and empirical studies (Davies, Cummings, & Winter, 2004; Kerig, 1995; Kretchmar & Jacobvitz, 2002).

Furthermore, family system theorists stress that family patterns must be contextualized, including consideration of developmental transitions and how these transitions may serve as informative catalysts for examining family functioning and children's development (e.g., Cowan, 1991; Minuchin, 1985). Thus, the present study examined how family functioning across multiple systems was associated with children's adjustment over the early school years. As a significant developmental milestone in the lives of children (Alexander & Entwisle, 1988; Cowan, Cowan, Ablow, Johnson, & Measelle, 2005; Perry & Weinstein, 1998; Vecchiotti, 2003), the early school years provide an informative window for examining how family relationship patterns may affect children's adaptation in extrafamilial contexts (Davies, Winter, & Cicchetti, 2006). For example, in navigating stage-salient tasks inherent in adapting to the school context such as establishing self-confidence, academic competence, peer relationships, and behavioral control in the school setting (e.g., Ladd & Burgess, 2001), children may turn to their family as a source of support and protection. Furthermore, numerous developmental theories suggest that children's internalizations of family experiences are used as blueprints for interpreting and responding to novel challenges outside the family (Cassidy, Kirsch, Scolton, & Parke, 1996; Greenberg, Speltz, & DeKlyen, 1993). Thus, dysfunctional family patterns may undermine children's successful adaptation in the school context across multiple domains of functioning.

Previous empirical work has reported linkages between various patterns of family functioning and children's adjustment. Cohesive families have been associated with lower internalizing and externalizing symptoms (e.g., Johnson, Cowan, & Cowan, 1999; Kerig, 1995; McHale & Rasmussen, 1998; Richmond & Stocker, 2006) and better peer relationships (e.g., McHale, Johnson, & Sinclair, 1999). In contrast, enmeshed family patterns have been related to elevated internalizing symptoms in older children and adolescents (Davies et al., 2004; Jacobvitz, Hazen, Curran, & Hitchens, 2004; Kerig, 1995) while disengaged family patterns have been linked with elevated externalizing problems (e.g., Davies et al., 2004; Kerig, 1995). However, despite the developmental significance of children's successful adaptation to school, scant research has examined relations between patterns in family functioning and longitudinal assessments of children's adjustment trajectories across multiple domains during the early school years. As the lone study to examine typologies of family functioning and children's adjustment during the early school years, Johnson (2003) identified three family typologies at kindergarten—cohesive, separate, and enmeshed family patterns—which in turn differentially predicted children's externalizing behaviors in the school context.

The present study seeks to build on the findings reported by Johnson (2003) in several important ways. First, from a methodological standpoint, Johnson utilized a triadic family interaction task in deriving typologies. While this assessment method provides important information regarding the complexity of mother–father–child interactions at a triadic level, the incorporation of assessments of multiple family contexts allows for a richer characterization of family processes. Accordingly, the present study also utilized assessments of family functioning at dyadic and triadic levels, including the interparental relationship, the parent–child dyadic relationship, and the mother–father–child relationship. The inclusion of distinct assessments of multiple systems allows for a comprehensive examination of patterns of functioning in multiple contexts. Second, from a developmental standpoint, Johnson (2003) restricted assessments of children's adjustment to examining changes in externalizing symptoms as a function of family patterns. Progress in understanding how different family patterns may eventuate in children's adjustment difficulties depends upon expanding beyond the usual focus on global assessments of children's psychological symptoms as outcome measures. To this end, the present study simultaneously examined theoretically defined domains of school adjustment (Ladd, 2003), which advocate for examining children's perceptions and appraisals of the school environment, involvement and conformity to classroom activities, and psychological reactions in the school context (e.g., internalizing and externalizing dimensions). Following previous operationalizations of domains of school adjustment, difficulties in emotional adjustment to school were defined as children's perceptions of loneliness, disengagement, and difficulties with peers in the school context (e.g., Buhs & Ladd, 2001), while difficulties in classroom engagement included children's problems in cooperating and actively participating within the context of classroom activities and rules (e.g., Birch & Ladd, 1997; Buhs & Ladd, 2001). Psychological reactions were assessed using internalizing and externalizing symptomatology.

In accordance with predictions derived from family systems theory, we hypothesized that enmeshed family patterns would be more strongly linked with internalizing symptomatology and difficulties in emotional adjustment to school. According to the theory (e.g., Minuchin, 1985), family enmeshment may engender greater internalizing difficulties in the school context through impinging developmental processes at a social-cognitive level (e.g., developing appraisals of the self, especially self-esteem and self-efficacy in external contexts) and social-emotional level (e.g., impinging on autonomy). Drawing from notions in the clinical literature that postulate that children in disengaged families may have a skewed sense of independence and autonomy, lack feelings of loyalty and belonging, and a diminished capacity for interdependence in social situations (e.g., Minuchin, 1974), we also hypothesized that disengaged family patterns would be most strongly linked to externalizing symptoms and difficulties in classroom engagement.

Method

Participants

The data for this study were drawn from a larger project designed to assess linkages between family relationships and children's coping and adjustment (Davies, Sturge-Apple, Winter, Cummings, & Farrell, 2006). The original pool of participants consisted of 236 mothers, fathers, and children. Families were included in the project if they met the eligibility criteria that the family had a child in kindergarten and had lived together for at least 3 years. The retention rate from across the three measurement occasions was 91%, resulting in a sample of 214 families (114 girls and 100 boys). Minimal differences were found between the sample who participated at all three waves and those participants who experienced attrition at any point in time over the two waves ($n = 22$). Therefore, in order to maximize study power, we elected to retain the full sample for study analyses and utilized full-information likelihood when estimating missing data within model analyses. Full-information maximum likelihood (FIML) is a widely accepted and advantageous technique for the imputation of missing data (see Enders, 2001). Two families were identified as multivariate outliers based upon Mahalanobis distance estimates greater than $p = .001$ *derived* from constructs specified in this study and were therefore excluded from the analyses (Arbuckle, 1997). Thus, the final sample for this study consisted of 234 mothers, fathers, and children (129 girls and 105 boys; mean age = 6.0 years, *SD* = 0.50 at Wave 1).

A primary goal of our recruitment strategies was to obtain a sample of families who, on the whole, exhibited diverse levels of interparental adversity and child vulnerabilities. Consistent with this goal, families were socioeconomically and demographically representative of the counties from which they were drawn. Median family income of the participants fell between $40,000 and $54,000, with 12% of the sample reporting household income below $23,000. A large proportion of the sample was European American (78.6%), followed by smaller percentages of African American (15.2%), Latin American (3.1%), Asian American (1.2%), Native American (0.2%), and other racial (1.7%) families. Families in the study also experienced a relatively wide range of adversity in the family as 33% of mothers and 43% of fathers in our sample could be classified as maritally dissatisfied based on scores below 100 on the Short Marital Adjustment Test (SMAT), with 53% of the couples containing at least one maritally dissatisfied partner (e.g., SMAT; Locke & Wallace, 1959). Further illustrating the occurrence of risk factors in our sample, proportions of children who scored in the clinical range ($t \geq 60$) on parent Child Behavior Checklist reports of internalizing and externalizing symptoms averaged 21% and 18%, respectively, a percentage that is modestly higher than a representative sample of U.S. children (Achenbach, 1991).

Procedures

Families visited the laboratories each year for two visits which were spaced one week apart. During the first visit at Wave 1, mothers, fathers, and the target child participated in a triadic interaction task in which members were instructed to play Jenga, an interactive game, for 15 min. The task was structured to elicit family involvement, competition, and interest. In addition, mothers and fathers participated in an interparental interaction task in which they were asked to manage and resolve two topics that resulted in common, but intense interparental disagreements (see Du Rocher Schudlich, Papp, & Cummings, 2004). Finally, assessments of dyadic parent–child relationships were obtained from separate 10-min play and cleanup tasks during the Wave 1 visit. The task consisted of 5 min of free play between the parent and child followed by a 5-min period in which the parent attempted to have the child clean up the toys. Fathers participated at the end of the first visit, while the mothers participated in the task at the end of the second visit. Videotaped records of these interactions were obtained for later coding of behaviors. Children's classroom teachers independently completed questionnaires at Waves 1, 2, and 3, spaced 1 year apart, to assess children's socioemotional adjustment in the school context.

Measures

Interparental relationship (Wave 1). Interaction behaviors of mothers and fathers during the interparental discussion tasks during the Wave 1 visit were evaluated using subscales from the System for Coding Interactions in Dyads (SCID; Malik & Lindahl, 1996). Molar ratings were yielded for each of the two interparental interactions on 5-point continuous scales ranging from 1 (*very low*) to 5 (*high*). These scales have been widely used across various populations with demonstrated reliability and validity in relation to similar interparental interaction measures (Malik & Lindahl, 1996).

Maternal and paternal hostility during each of the interparental interactions were assessed using the following SCID scales: (a) Verbal Aggression, which is defined as the level of hostile and aggressive behaviors and verbalizations each partner expresses toward the other, and (b) Negativity and Conflict, which reflects spousal displays of anger, frustration, and tension. Intraclass correlation coefficients, which indexed the reliability of two independent coders for 25% of the interactions, ranged from .84 to .97 for mothers and fathers across each of the interactions. Husband and wife behaviors across the two interactions were then summed to create one score for each of the three behaviors reflecting the degree to which each couple used a specific strategy. Coders also rated each couple as a whole using the Negative Escalation code from the SCID, which assesses tendencies of the partners to reciprocate and escalate displays of negative affect and hostility. The intraclass correlation coefficients, which reflect the interrater reliability of two independent coders for 25% of the interactions, were .96 and .97 across the two interactions. Principal components analysis of the individual and couple ratings indicated a single factor, accounting for 91% of the variance in scores. In addition, internal consistency in this sample was .79 for the ratings. Thus, scales were aggregated to form an interparental hostility composite score.

Consistent with the interparental hostility measurement battery, mother, father, and couple withdrawal during the interparental interactions were assessed using SCID scales (Malik & Lindahl, 1996). Maternal and paternal withdrawal were specifically assessed through the SCID Withdrawal scale, with higher scores reflecting displays of repeated, prolonged, and tense forms of detachment and avoidance during the interparental interactions. Intraclass correlation coefficients, which index interrater reliability, ranged from .85 to .96 for mothers and fathers across each of the two interactions. To obtain a third index of withdrawal, coders rated the couple as a whole using the Pursuit/Withdraw Scale from the SCID, which assesses the extent to which one partner attempts to avoid discussion while the other partner attempts to attempts to continue the discussion. The intraclass correlation coefficients indexing interrater reliability for the couple were .90 and .91 across the two interactions. Principal components analysis of the individual and couple ratings indicated a single factor, accounting for 90% of the variance in scores. In addition, internal consistency in this sample was .76 for the ratings. Thus, scales were aggregated to form an interparental withdrawal composite score.

Parent–child relationship (Wave 1). Interactions between mother–child and father–child dyads and were assessed during the play and cleanup tasks at Wave 1 using a global rating scale assessing the frequency, quality, and intensity of behaviors on a 9-point scale, ranging from 1 (*not at all characteristic*) to 9 (*mainly characteristic*).

Parent behaviors were evaluated using subscales from the Iowa Family Interaction Rating Scales (IFIRS; Melby & Conger, 2001). The widely used IFIRS has demonstrated validity and reliability across studies (see Melby & Conger, 2001). Coders, who were blind to ratings of the interparental and triadic interactions, independently rated mother and father parenting behaviors. Parental emotional availability during the parent–child interaction task was assessed by the Warmth/Support, Positive Reinforcement, and Neglecting/Distancing scales of the IFIRS. The Neglecting/Distancing scale measures parental indifference, apathy, and unresponsiveness during the parent–child interactions, whereas Warmth/Support is characterized by displays of affection and support toward the child. Finally, Positive Teinforcement reflects positive parental responses (e.g., praise, approval, rewards) to appropriate or exemplary child behavior. Separate ratings of the codes were obtained during play and compliance components of the interaction task. Intraclass correlation coefficients, which reflected interrater reliability among two coders for 21% of the interactions, ranged from .91 and .94 for ratings of the parenting codes for mothers and fathers. Principal components analysis of mother and father ratings indicated a single factor, accounting for 64% of the variance in scores, while the internal consistency coefficient in this sample was .83. Thus, ratings were aggregated to form a single emotional availability score (the Neglecting/Distancing scales were reverse coded prior to compositing).

Parental intrusiveness was assessed with the Intrusiveness subscale of the IFIRS system, measuring intrusive and overcontrolling behaviors by the parents. In structured tasks, this behavior may be manifested by extreme concern about completing the task, constriction of the child's autonomy and hindering the child from exploring and setting the pace for the task. Intraclass correlation coefficients ranged from .82 and .88 for mothers' and fathers' ratings, respectively. Principal components analysis of mother and father ratings indicated a single factor, accounting for 53% of the variance in scores, and internal consistency in this sample was .71. Thus, ratings were composited to form a single intrusiveness score.

Child behaviors were evaluated using the Relatedness scale that is designed to assess children's comfort and positive affect in parent–child interactions (Davies, 2002). The Relatedness subscale captures the degree to which a child's interest in initiating and maintaining pleasurable and harmonious interactions with the parent as a means of promoting positive affect and well-being in the dyad. Ratings on the 5-point scale take into account the number of attempts by the child to engage the parent, the duration of the attempts, and the quality of the accompanying affect. Whereas low scores reflects children's attempts to engage the parent that are brief, mild, and relatively superficial or ambiguous in intent, higher scores are characterized by repeated, socially skilled, prolonged, and enthusiastic attempts to initiate or maintain positive relations with the parent. Intraclass correlation coefficients for interrater reliability ranged from .70 to .84 for children's ratings with mothers and fathers ratings, respectively. Principal components analysis of ratings for children across tasks indicated a single factor, accounting for 84% of the variance in scores, and internal consistency in this sample was .83. Thus, ratings were composited to form a single relatedness score.

Mother–father–child relationship (Wave 1). Family-level behavior was evaluated using constructs from the Coparenting and Family Rating system (CFRS; McHale, Keursten-Hogan, & Lauretti, 2001) and the System for Coding Interactions and Family Functioning (SCIFF; Lindahl & Malik, 1996). The global rating scales of these systems were coded on a 5-point scale, ranging from 1 (*very low*) to 5 (*excessive*). Specifically, in order to index coparenting difficulties in the presence of the child, the Competition and Cooperation subscales from the CFRS were utilized. The Competition subscale captures active parental intrusiveness and interference in the other parent's child-directed activities, while the Cooperation subscale measures the ability of coparents to work as a team and support one another in family interactions. Finally, the Cohesiveness subscale of the SCIFF was utilized to index the sense of unity, togetherness, and closeness within the triad. Intraclass correlation coefficients, which reflected interrater reliability among two coders for 25% of the interactions, ranged from .60 to .73 at Wave 1 across the three scales.

Child internalizing and externalizing symptoms (Waves 1, 2, and 3). Internalizing and externalizing symptoms were assessed using their respective broadband scores from the Teacher Report Form (Achenbach, 1991). For each item, teachers indicated whether or not each statement is true of the child on a 3-point ordinal scale, ranging from 0 (*not true*) to 2 (*very true*). Broadband scores for internalizing symptoms included the Withdrawn/Depressed, Somatic Complaints, and Anxious/Depressed subscales. Internal consistencies for internalizing scores across the three waves ranged from .83 to .90. Broadband externalizing symptoms were based on the Aggression and Delinquent subscales. Internal consistencies for the broadband externalizing score across the three waves ranged from .92 to .95.

Children's difficulties in classroom engagement (Waves 1, 2, and 3). To obtain an assessment of children's difficulties with engagement in the classroom, two scales were used. First, teacher report on the Hyperactivity/Distractibility subscale of the Child Behavior Scale (CBS) was used as a measure of children's behavior at school (Ladd & Profilet, 1996). The subscale consists of four items (e.g., ''poor concentration, attention span'') rated along a 3-point scale ranging from 1 (*doesn't apply*) to 3 (*certainly applies*). Internal consistency in this sample ranged from .84 to .87 across the different time points. Prior research has supported the validity of the CBS (Ladd & Profilet, 1996). Second, teachers reported on the Cooperative Participation subscale from the Teacher Rating Scale of School Adjustment (TRSSA; Birch & Ladd, 1997). The subscale consists of seven statements assessing teacher appraisals of children's participation in the classroom (e.g., ''follows teacher's directions'' and ''accepts responsibly for a given task'') on a 3-point scale from 1 (*doesn't apply*) to 3 (*certainly applies*). The scale was later reversed such that higher scores reflected lower levels of cooperative participation. The TRSSA has demonstrated reliability as well as content and construct validity (Birch & Ladd, 1997). Internal consistency across the three waves ranged from .76 to .90. To create an overall Children's Classroom Difficulties scale for each time point, the two subscales were summed to create a combined score. Internal consistency of the overall composite exceeded .85 at each time point.

Children's difficulties in emotional adjustment to school (Waves 1, 2, and 3). To obtain an assessment of children's difficulties in emotional adjustment to school we followed previous conceptualizations of poor emotional adjustment in the school as reflecting children's experiences of loneliness, disengagement, and difficulties with peers in the school context (e.g., Buhs & Ladd, 2001). Three subscales, rated on a 3-point scale from 1 (*doesn't apply*) to 3 (*certainly applies*), were used to assess this construct. First, the School Avoidance and School Liking subscales from the TRSSA (Birch & Ladd, 1997) were used. The school avoidance scale consists of five items that assess teacher's perceptions of a child's efforts to avoid the classroom (e.g., ''makes up reasons to go home from school, feigns illness at school''). Internal consistencies in this sample were .70 at Time 1, .77 at Time 2, and .83 at Time 3. Second, the School Liking subscale is a five-item index of teacher's perceptions of how much a child likes being in school (e.g., ''likes to come to school, has fun at school''). Internal consistencies in this sample were .86 at Time 1, .86 at Time 2, and .88 at Time 3. This scale was reverse coded. To obtain an assessment of children's ability to connect with peers in school, teacher report on the Prosocial Behavior subscale of the CBS was used (Ladd & Profilet, 1996). This scale assesses children's tendencies to be empathic, cooperative, and self-sacrificing with peers and was measured with items that referred to behaviors such as helping, concern, and kindness. Internal consistency in this sample ranged from .81 to .93 across the different time points. Items were reverse coded such that higher scores reflect poorer functioning with peers. To create an overall Children's Emotional Adjustment Difficulties scale for each time point, the three subscales were summed to create a combined score. Internal consistency of the overall composite exceeded .81 at each time point.

Results

The results are presented in two sections. The first details the results of the latent class analysis of the family interaction constructs and the second presents latent growth curve (LGC) analyses of children's adjustment trajectories in the school context as a function of family group. All models were estimated using the latent variable software Mplus, Version 5.1 (Muthén & Muthén, 2006).

Examination of Family Typology at Wave 1

Latent Profile Analysis (LPA; e.g., Hagenaars & McCutcheon, 2002), was utilized to determine family profiles. LPA is a variant of latent class analysis techniques and is used when manifest indicators of the categorical latent variable are on a continuous scale as family assessments were in the present study. As in other latent variable modeling techniques, the profiles in LPA are not observable directly from the data but are identified based on a function of a set of the continuous manifest indicators. To determine the optimal number of profiles, we used two likelihood tests for comparing nested models including the Bayesian information criterion (BIC; Schwartz, 1978) and the Vuong–Lo–Mendell–Rubin likelihood ratio test (VLMR; Lo, Mendell, & Rubin, 2001). The model with the lowest BIC value is generally considered the best-fitting model. As a further test of model fit, the VLMR provides a p value that indicates whether the k-1 profile model is rejected in favor of the k profile model (see Nylund, Asparouhov, & Muthen, 2007, for more details). Finally, it is also important to consider the practical applications of profile solutions along with statistical measures of model fit. Therefore, substantive implications were also taken into account when determining optimal number of profiles (e.g., Muthén, 2004). Furthermore, for each family, the LPA analysis produces posterior (Bayesian) probabilities of likelihood of membership in each latent profile, and families are classified into the latent profile for which the posterior probability is highest. The probabilities are a function of the model's parameters (estimated conditional response probabilities and estimated prevalence of each latent class). Inspection of these classifications and posterior probabilities is necessary to ensure that families have a high likelihood of being classified correctly. Finally, the latent profile variable indicators (e.g., interparental withdrawal, interparental hostility, parental warmth, parental intrusiveness) were each standardized to aid in interpretation of the latent profiles.

To examine the best profile solution, we extracted two, three, and four latent profiles from our models. Results of the LPA revealed that model fit was best for the three-profile model (BIC = 5897.54; VLMR 2 vs. 3 model = 108.61, $p < .05$), relative to the two (BIC = 5953.89; VLMR 1 vs. 2 model = 163.56, $p < .05$) and four (BIC = 5889.39; VLMR 3 vs. 4 model = 61.88, $p = .49$, *ns*) profile models. Thus, while the difference in BIC values between the three-profile and four-profile models was positive ($\Delta = 8.15$), the VLMR indicated that the three-profile solution was a significantly better fit than a two-profile solution, and the VLMR indicated that the four-profile solution did not provide a statistically improved fit relative to the three-profile solution. Therefore, the three-profile solution is regarded as a statistically comparable but more parsimonious model than the four-profile solution. In terms of substantive considerations, the three-profile solution cohered into cohesive, disengaged, and enmeshed patterns among indicator variables. Although the four-group solution contained the three profiles as culled in the three-group solution, it contained one profile with only four families that evidenced high levels of interparental hostility, moderate levels of parental emotional availability and intrusiveness, and low levels of withdrawal. Accordingly, this small group of families closely resembled the profile of an enmeshed pattern identified in the three-profile solution. Given the results from the VLMR, conceptual overlap of the profiles between the three- and four-profile solutions, as well as the extremely small sample size in the fourth profile, we retained the three-profile solution.

Table 1 displays the probabilities of the three-profile LPA model and reveals the degree of fit among the profiles for the families in the study. In the present study, average latent class probabilities for most likely latent class membership by latent class were excellent. Specifically, families who were profiled as disengaged had a 90.6% average probability of fitting the disengaged profile, a 6.5% probability of fitting the cohesive profile, and a 2.9% probability of fitting the enmeshed profile. Families who were profiled as cohesive had a 95.4% average probability of fitting the cohesive profile, a 1.9% probability of fitting the disengaged profile, and a 2.7% probability of fitting the enmeshed profile.

Table 1
Latent Profile Probabilities for Most Likely Class Membership by Latent Profile Based on Estimated Model

	Latent Profile		
	Disengaged	Cohesive	Enmeshed
Latent profile classification			
Disengaged	90.6%	6.5%	2.9%
Cohesive	1.9%	95.4%	2.7%
Enmeshed	4.5%	4.1%	91.4%

Finally, families who were classified as enmeshed had a 91.4% average probability of fitting the enmeshed profile, a 4.5% probability of fitting the disengaged profile, and a 4.1% probability of fitting the cohesive profile. Inspection of the cases revealed that three families displayed some indiscrimination among profiles. Their highest profile probability (range = .49–.52) coupled with high cross-profile probabilities (range = .42–.45) among the other profiles suggested that they could not be validly classified within one of the three latent profiles (e.g., Nagin, 2005). Therefore, these families were not retained in the analyses.

Table 2 shows the means and standard deviations of the standardized measures of family process for each of the three family profiles identified by the latent profile analyses as well as the results of analysis of variance comparisons among the groups. The first profile displayed low levels of interparental hostility and withdrawal, parental intrusiveness, and family competition as well as high levels of parental emotional availability, child relatedness, and family cooperation and cohesiveness. Based on earlier empirical work and theoretical conceptualizations, these families were classified as exhibiting a cohesive pattern of interactional style within family subgroups. Families in the second profile displayed very high levels of interparental hostility, parental emotional availability and intrusiveness, and family competition along with lower levels of interparental withdrawal, child relatedness, and family cooperation and cohesiveness. Given evidence of high levels of hostility and intrusiveness coupled with modest to moderate levels of warm and supportive parenting, we classified these families as enmeshed. Finally, the third profile of families was classified as displaying disengaged patterns of interaction. These families were characterized by relatively high levels of interparental withdrawal, parental intrusiveness, while also exhibiting low levels of interparental hostility, parental emotional availability, family cooperation, competition and cohesiveness, and child relatedness.

In addition, several family background variables were included in the analyses to examine their relation to family typology membership. Child gender, maternal age, and family income were specified as covariates in our latent profile analyses to incorporate their effects on the odds of model classification into profiles. Using the cohesive profile as the comparison group, two comparisons were made for each covariate: (a) the likelihood of being categorized as an enmeshed family compared to a cohesive family and (b) the likelihood of being

Table 2
Means, Standard Deviations, and ANOVA Comparisons of the Three Family Typologies on Seven Defining Variables

	Cohesive (C; *n* = 137)		Enmeshed (E; *n* = 51)		Disengaged (D; *n* = 43)			
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*(2, 230)	Post hoc
Wave 1								
Interparental hostility	−.46	.53	1.47	.79	−.27	.64	187.50***	E > C, D
Interparental withdrawal	−.36	.67	−.18	.74	1.38	.97	90.50***	D > E, C
Parental emotional availability	.31	.82	.01	1.02	−.99	.86	36.17***	E, C > D
Parental intrusiveness	−.14	.95	.09	1.07	.34	.99	4.20***	D > C
Child relatedness	.18	.96	−.12	1.01	−.44	.98	7.23***	E, D > C
Triadic competition	−.08	.90	.40	1.23	−.28	.88	6.20***	E > C, D
Triadic cooperation	.18	.91	−.16	.98	−.37	1.18	5.97***	C > D, E
Triadic cohesiveness	.27	.95	−.20	.92	−.61	.91	15.59***	C > E, D

Note. Post hoc comparisons used Tukey's HSD to control for alpha level, ">" refers to significantly larger whereas "," refers to not significantly different at alpha = .05 level. ANOVA = analysis of variance.
$^{***}p \leq .001$.

categorized as a disengaged family compared with a cohesive family. Results indicated that only family income had an effect on probability of class membership for disengaged versus cohesive families. The logistic regression coefficient for family income (–.89, $p < .05$) indicated that as family income increased the probability of being classified as disengaged versus cohesive decreased. Other class comparisons were not significant.

Links Between Family Typology and Children's Adjustment Trajectories

Our final set of analyses explored whether children's adjustment trajectories were associated with family typology using Mplus 5.1 statistical software for estimating trajectories for each group (Muthén & Muthén, 2006). Accordingly, unconditional LGC models were specified for each dependent measure (e.g., internalizing symptoms) within each family profile (e.g., cohesive). Unconditional LGC allow for the examination of the existence and significance of average intercepts and slopes within each family profile. Because our analyses involved both categorical (family profile) and continuous (outcome) measures, we utilized the MPlus options for estimating mixture models. First, we examined whether the slope of each family typology was significantly different from zero. Second, we examined whether the slopes of each family typology are significantly different from each other. To accomplish this second goal, we tested for significant differences among the family profiles on model-derived intercepts and slopes. However, because a chi-square test is not provided in mixture models, the chi-square difference test is computed as $2d$ where d is the difference between the log-likelihood values from the two models being compared (McLachlan & Peel, 2000). The difference in degrees of freedom is computed as the difference in the number of parameters.

The LGC model of children's linear externalizing trajectories was examined first. Mean intercept values were positive and significant in each of the three groups ($p < .05$). Log-likelihood comparisons among the different family typologies revealed that children in disengaged families had significantly higher intercept values when compared with children in enmeshed and cohesive families (see Table 3). The means of the slope factors for cohesive, disengaged, and enmeshed families were 0.24 ($p = 0.13$), 1.37 ($p < .05$), and .15 ($p = 0.50$), respectively (see Figure 1). These findings indicated that children's externalizing symptoms in the sample increased across all family profiles; however, the magnitude of the increase was only significant for children from disengaged families. Furthermore, log-likelihood comparisons among the family profiles on linear slope means revealed that children from disengaged families had significantly

Table 3
Results From Log-Likelihood Difference Tests for Free to Vary Model Versus Constrained Models

Model	log*L*	*df*	Δlog*L*	Δ*df*
Model: Children's externalizing symptoms trajectories				
Overall free to vary model	2,015.57	14	—	—
Intercept comparisons				
Disengaged = Cohesive	2,019.05	13	6.96*	1
Disengaged = Enmeshed	2,018.39	13	5.64*	1
Cohesive = Enmeshed	2,015.58	13	0.02	1
Slope comparisons				
Disengaged = Cohesive	2,019.98	13	8.82*	1
Disengaged = Enmeshed	2,019.24	13	7.34*	1
Cohesive = Enmeshed	2,015.60	13	0.06	1
Model: Children's internalizing symptoms trajectories				
Overall free to vary model	1,712.26	14	—	—
Intercept comparisons				
Disengaged = Cohesive	1,712.49	13	0.46	1
Disengaged = Enmeshed	1,712.33	13	0.14	1
Cohesive = Enmeshed	1,712.28	13	0.04	1
Slope comparisons				
Disengaged = Cohesive	1,715.86	13	7.20*	1
Disengaged = Enmeshed	1,712.45	13	0.38	1
Cohesive = Enmeshed	1,714.50	13	4.48*	1
Model: Children's difficulties in classroom engagement trajectories				
Overall free to vary model	1,599.87	14	—	—
Intercept comparisons				
Disengaged = Cohesive	1,603.84	13	7.94*	1
Disengaged = Enmeshed	1,602.56	13	5.38*	1
Cohesive = Enmeshed	1,599.88	13	0.02	1
Slope comparisons				
Disengaged = Cohesive	1,602.04	13	4.34*	1
Disengaged = Enmeshed	1,601.87	13	4.00*	1
Cohesive = Enmeshed	1,599.90	13	0.06	1
Model: Children's difficulties in emotional adjustment to school trajectories				
Overall free to vary model	1,646.99	14	—	—
Intercept comparisons				
Disengaged = Cohesive	1,648.07	13	2.16	1
Disengaged = Enmeshed	1,648.17	13	2.36	1
Cohesive = Enmeshed	1,647.08	13	0.18	1
Slope comparisons				
Disengaged = Cohesive	1,649.70	13	5.42*	1
Disengaged = Enmeshed	1,647.07	13	0.16	1
Cohesive = Enmeshed	1,648.91	13	4.02*	1

Note. Significance level for 2Δlog*L* tests is based on chi-square distribution.
*$p < .05$.

accelerated externalizing trajectories in comparison to children from cohesive and enmeshed families. Average slopes for children from disengaged and enmeshed families were no different suggesting similar change over time.

The LGC model of children's linear internalizing trajectories was examined next (see Figure 2). Mean intercept values were positive and significant in each of the three groups ($p < .05$). Log-likelihood comparisons among the different family typologies revealed no significant differences in average intercept values, suggesting that children started at relatively similar levels of internalizing symptoms (see Table 3). Findings also revealed that children from disengaged families had a significant and positive mean slope ($\mu = .74$, $p < .05$), children from enmeshed families had a significant and positive mean slope ($\mu = .55$, $p < .05$), and children from cohesive families had a nonsignificant and positive mean slope ($\mu = .05$, $p = .63$). Furthermore, average slopes for children from disengaged and enmeshed families were no different suggesting similar change over time.

The LGC model of children's linear trajectories of classroom difficulties revealed that while mean intercept values were positive and significant across

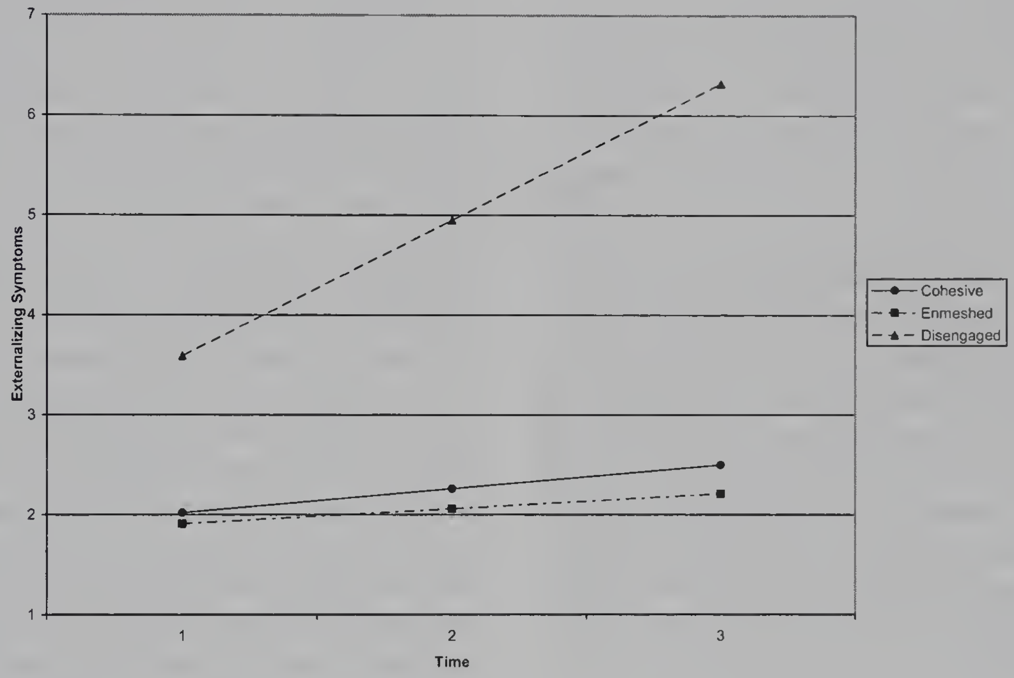


Figure 1. Trajectories for teacher-reported externalizing symptoms.

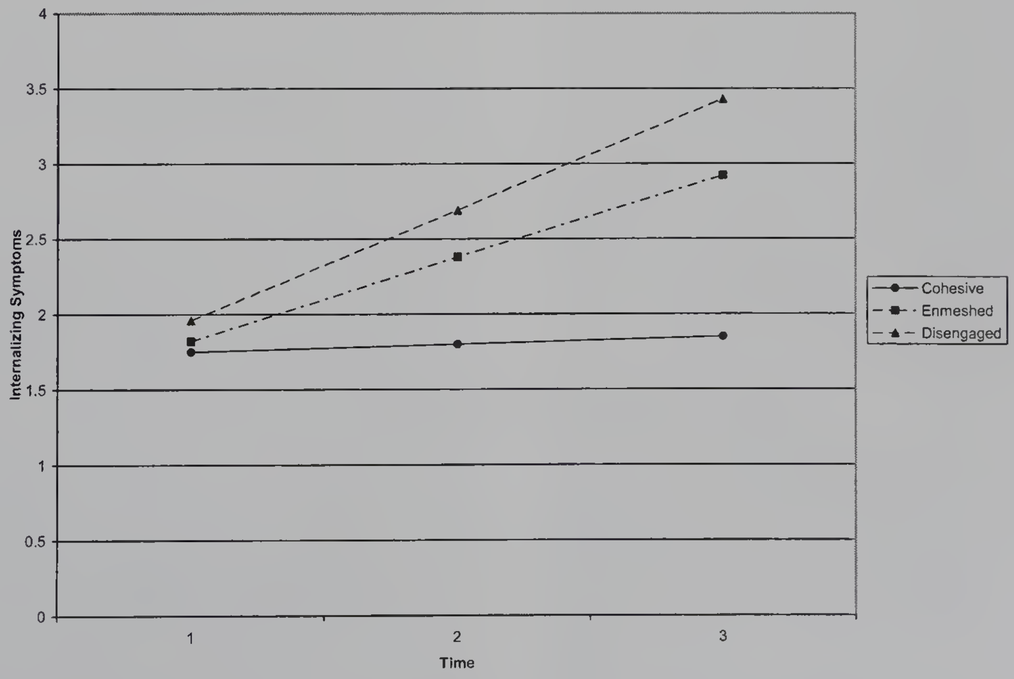


Figure 2. Trajectories for teacher-reported internalizing symptoms.

the three groups, children in disengaged families had significantly higher intercept values when compared to children in enmeshed and cohesive families (see Table 3). Furthermore, children from disengaged families had a significant and positive mean slope ($\mu = .58$, $p < .05$) while children from cohesive ($\mu = .18$, $p = .07$) and enmeshed ($\mu = .22$, $p = .09$) families had nonsignificant slopes (see Figure 3). Log-likelihood comparisons among the means further revealed that children from disengaged families had elevated trajectories of classroom difficulties in comparison to children from enmeshed and cohesive families. No other differences were detected.

Our final analysis examined the LGC model of children's difficulties in emotional adjustment to school. Findings revealed that no differences among the groups on the average intercept values (Table 3); however, children from both disengaged and enmeshed families had significant and positive mean slopes (μs = .57 and .45, respectively, $p < .05$) while children from cohesive families had a nonsignificant slope ($\mu = .03$, $p = .81$). Thus, children from disengaged and enmeshed families experienced significant increases in difficulties with emotional adjustment to school over time. Log-likelihood comparisons among the means further revealed that children from enmeshed and disengaged families had elevated trajectories of classroom emotional difficulties in comparison with children from cohesive families only (see Figure 4). No other differences were detected.

Discussion

The findings of the present study demonstrate the value of utilizing a latent class analysis of family interaction constructs for delineating distinct configurations of relationship and interactional parameters in family systems. In addition, patterns of family interactions had developmental utility, with distinct typologies meaningfully associated with differences in longitudinal trajectories of child adjustment in the school context (e.g., Jacobvitz et al., 2004; Johnson, 2003; Kerig, 1995; McHale et al., 2004). Thus, the present study highlights new avenues for the analysis of family functioning and implications for child development in the school context. From a substantive perspective, the present study adds to our understanding of the developmental implications of different configurations of family functioning proposed in family systems theory. From a methodological perspective, the specific focus of coupling latent class analysis with LGC modeling illuminates a new way of examining complex family systems conceptualizations within a developmental framework.

Consistent with past empirical and clinical work, three specific family typologies could be reliably identified utilizing multilevel assessments of family interactions spanning both dyadic and triadic systems. Solutions extracted a cohesive profile across the interparental, parent–child, and triadic subsystems in accordance with previous classifications of families who effectively maintain relationship

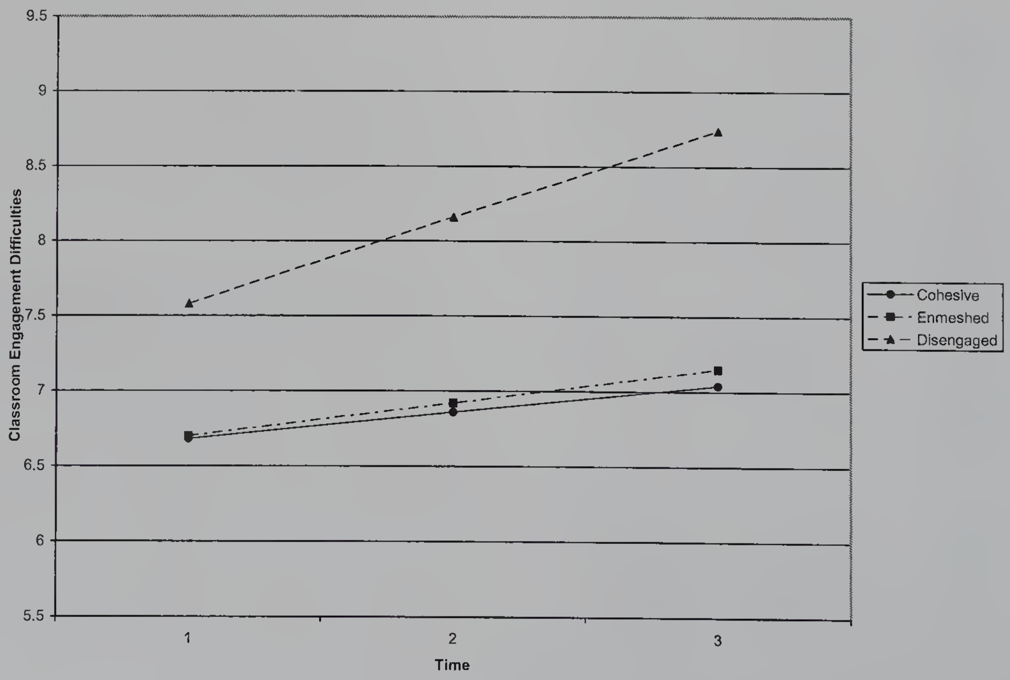


Figure 3. Trajectories for teacher-reported difficulties in classroom engagement.

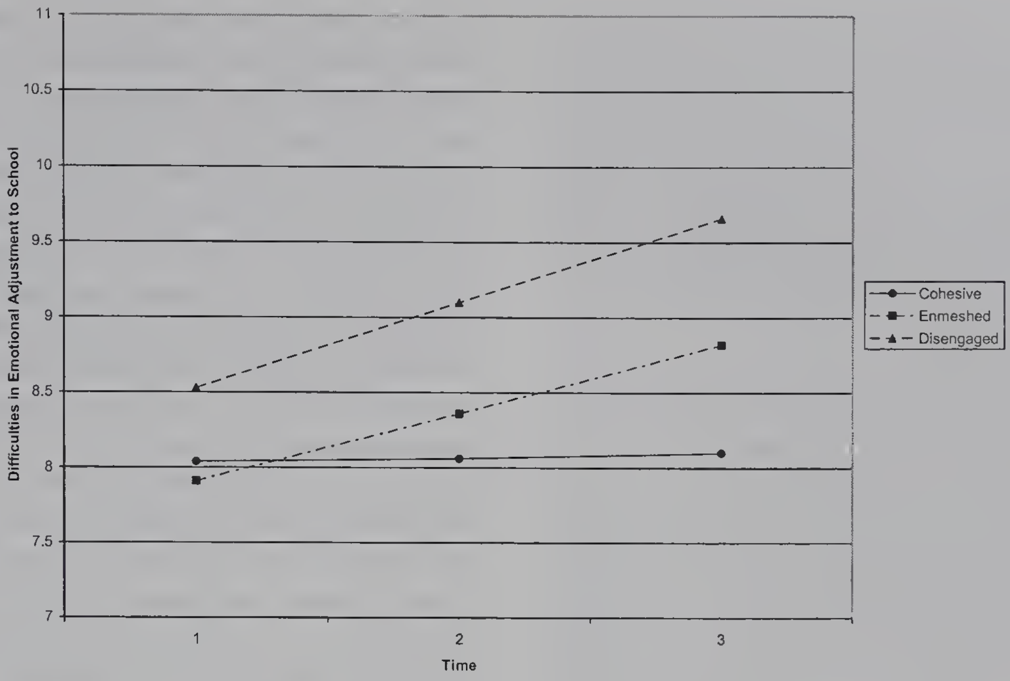


Figure 4. Trajectories for teacher-reported difficulties in emotional adjustment to school.

boundaries by permitting members to access resources without threatening their autonomy (Davies et al., 2004; Kerig, 1995). Families classified as disengaged exhibited low levels of positive affect across family systems along with high levels of diminished engagement and intrusiveness. Enmeshed families evidenced high levels of hostility and negative affect, intrusiveness, and competitiveness; low levels of cohesion; and modest levels of warmth and availability in parent–child relationships. Difficulties in regulating negative affect and maintaining positive affect in the presence of diffuse boundaries map onto previous distinctions in the literature. Thus, this study adds to a growing line of research (e.g., Davies et al., 2004; Johnson, 2003; O'Connor et al., 1998) that advocates for incorporating pattern-based approaches in studying family relationships across multiple family systems.

To address the developmental utility of these specific configurations, we examined how patterns of family functioning were associated in meaningful ways with changes in children's adjustment in the school context. As hypothesized, children in enmeshed families exhibited significantly greater slopes of internalizing symptomatology as well as difficulties navigating emotional adjustment to school than children from cohesive groups. While this pattern of findings requires replication before drawing any definitive conclusions about the developmental timing of enmeshed family processes for children, it does raise some plausible interpretations. First, children who have experienced erosion of family boundaries inherent in enmeshed families may find themselves emotionally drawn into family difficulties, which in turn may amplify worries and distress about family relationships (e.g., Davies et al., 2004). In turn, this preoccupation and concern about family problems may detract from the demands of the school context and result in greater emotional disengagement from school. In addition, highly enmeshed subsystems increase the likelihood of various constellations of relationships within the family that are highly toxic to children's development (e.g., Jacobvitz et al., 2004). For example, enmeshed families may restrict children's autonomy by promoting dysfunctional family roles characterized by parent–child alliances, spousification of children (i.e., the child serves as a primary source of intimacy for the parent), or parentification of children (i.e., the child serves as a caregiver to the parent; Kerig, 1995; Sroufe, Jacobvitz, Mangelsdorf, DeAngelo, & Ward, 1985). Thus, increased exposure to interparental negative affect, integration into inappropriate family constellations, and lowered autonomy may undermine children's successful adaptation in extrafamilial settings such as the school context.

Moreover, given that children from enmeshed families were comparable to cohesive children in their initial levels of internalizing symptoms, it appears that the early elementary school years may be a sensitive period for the role of enmeshed family processes in the development of internalizing difficulties (e.g., Cicchetti & Toth, 1998). Children's adaptation to school is associated with significant cognitive changes including growing abilities for abstract representations and reasoning ability as well as developmental shifts in the complexity of self-cognitions, with a significant transition occurring in the 7- to 8-year-old range (e.g., Harter, 1983). In turn, these advancing abilities over the early school years may potentiate children's vulnerability to internalizing symptoms when experiencing family enmeshment.

Consistent with previous studies examining links between disengaged family patterns and children's externalizing behaviors (e.g., Davies et al., 2004; Kerig, 1995), the present study found that disengaged family patterns were associated with higher initial values on externalizing symptomatology and difficulties in classroom engagement. In addition, examination of slope trajectories was also consistent with prior empirical work in the sense that children from disengaged families had significantly greater slopes for classroom engagement difficulties and externalizing problems relative to both cohesive and enmeshed family profiles. Thus, in the case of disengaged families, it may be that the family's failure to fulfill the responsibility of providing support, resources, and protection results in children's difficulties in regulating behavior in the school context. In attempting to understand the mechanisms by which these associations occur, it may be that it is adaptive for children to dismiss or downplay the value of interpersonal relationships when confronted with fragmented and emotionally detached interaction patterns in the family (e.g., Minuchin, 1974). Unfortunately, when these responses are exhibited within the context of the school, a maladaptive pattern may result, including greater use of aggressive and inappropriate behaviors, greater alienation from teachers, and greater inability to attend to and adhere to rules and regulations for school functioning.

Finally, an intriguing finding arising in the present study was that children in disengaged families were comparable to children in enmeshed families in exhibiting a heightened vulnerability to internalizing symptoms and difficulties in emotional adjustment to school. In the context of the prevailing notion that the rigid boundaries and emotional distance characteristic of disengaged families is a more potent prognosticator of externalizing difficulties, these findings may be somewhat surprising. However, these results are consistent with an emerging body of literature proposing cascading models in development. First, cascade models of family dynamics (e.g., Christensen & Heavey, 1990; Sturge-Apple et al., 2006) suggest that disengaged family processes may be particularly pathogenic mechanisms that might not only engender alienation and disenfranchisement in family members but may also promote significant doubts and worries in children about the ability of the family to serve as a source of protection in the face of external demands. Thus, the barren emotional climate present in a disengaged family network may amplify adaptive difficulties across several domains of functioning. In line with this prediction, developmental cascade models (e.g., Masten et al., 2005) theorize that a constellation of difficulties across domains may largely be the function of early occurring externalizing problems in school children. Specifically, the cascade model proposes that externalizing symptoms evoke negative responses by peers and school staff, which in turn sets in motion a cascade of processes that increase children's internalizing symptoms and school problems. The presence of significantly higher initial levels of externalizing and classroom engagement difficulties in children in disengaged families in the current study coupled with the significant changes in internalizing and emotional adjustment domains over time is in accordance with the progressive, transactional effects proposed by Masten et al. (2005).

Several limitations and future directions warrant discussion. First, the relatively short time period utilized limits our generalizability of the developmental trajectories identified. Future work incorporating longer assessment windows may shed additional light on the implications of these family typologies for children's adjustment over time. Second, future research should complement the longitudinal assessment of adjustment problems in the present study with comparable prospective assessment of family typologies to examine how stability and change in family typologies correspond with trajectories of adjustment over time. Third, our models in the present study did not incorporate children's coping in stressful family systems and might have served to further elucidate the links between family typologies and their adjustment. Fourth, minimal differences were found on average intercepts of children's internalizing symptoms as well as emotional adjustment to school. Given that the impact of family patterns on children's adjustment in the school context may take some time to evolve, particularly with respect to internalizing and emotional domains, the lack of differences in the intercept is not surprising. However, because this is one of the first articles to chart these relations during this developmental period, we exercise caution in interpreting these null findings. Finally, results are based on a community sample; the function, correlates, and sequelae of family configurations in high risk and clinical populations may be different from those reported here.

In summary, our results suggest that employing pattern-based approaches to family functioning and child development may advance our understanding of the developmental sequelae of children's adaptation within different typologies of family functioning. Consistent with models of developmental psychopathology, maladaptive family patterns may potentiate children's trajectories of adjustment difficulties during the early school years. However, the findings here suggest that future research should consider multiple domains of adaptation simultaneously in order to fully explore how family functioning impacts children across an array of adjustment outcomes. At a broader level, this study also has implications for the utility of broadening transactional approaches to children's development in extrafamilial contexts as a function of different interactional patterns within the family system. Thus, delineating the interplay between family relationships may further advance understanding of child vulnerability and resilience within extrafamilial contexts.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist: 4-18 and 1991 profile.* Burlington: University of Vermont Department of Psychiatry.

Alexander, K. L., & Entwisle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53,* 2.

Arbuckle, J. L. (1997). *Amos user's guide version 3.6.* Chicago: SmallWaters.

Belsky, J., & Fearon, R. M. P. (2004). Exploring marriage–parenting typologies and their contextual antecedents and developmental sequelae. *Development and Psychopathology, 16,* 501–523.

Birch, S. H., & Ladd, G. W. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology, 35,* 65–79.

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology, 37,* 550–560.

Cassidy, J., Kirsch, S., Scolton, K., & Parke, R. D. (1996). Attachment and representations of peer relationships. *Developmental Psychology, 32,* 892–904.

Christensen, A., & Heavey, C. L. (1990). Gender and social structure in demand/withdraw pattern of marital conflict. *Journal of Personality and Social Psychology, 59,* 73–81.

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist, 53,* 221–241.

Cowan, P. A. (1991). Individual and family life transitions: A proposal for a new definition. In P. A. Cowan & E. M. Hetherington (Eds.), *Family transitions* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Cowan, P. A., & Cowan, C. P. (2002). Interventions as tests of family systems theories: Marital and family relationships in children's development and psychopathology. *Development and Psychopathology, 14,* 731–760.

Cowan, P., Cowan, C., Ablow, J., Johnson, V. K., & Measelle, J. (2005). *The family context of parenting in children's adaptation to elementary school.* Mahwah, NJ: Erlbaum.

Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology, 48,* 243–267.

Davies, P. T. (2002). *Emotional Security Scales.* Unpublished coding system, University of Rochester, Rochester, NY.

Davies, P. T., Cummings, E. M., & Winter, M. A. (2004). Pathways between profiles of family functioning, child security in the interparental subsystem, and child psychological problems. *Development and Psychopathology, 16,* 525–550.

Davies, P. T., Sturge-Apple, M. L., Winter, M. A., Cummings, E. M., & Farrell, D. (2006). Child adaptational development in contexts of interparental conflict over time. *Child Development, 77,* 218–233.

Davies, P. T., Winter, M. A., & Cicchetti, D. (2006). The implications of emotional security theory for understanding and treating childhood psychopathology. *Development and Psychopathology, 18,* 707–735.

Du Rocher Schudlich, T. D., Papp, L. M., & Cummings, E. M. (2004). Relations of husbands' and wives' dysphoria to marital conflict resolution strategies. *Journal of Family Psychology, 18,* 171–183.

Enders, C. K. (2001). A primer on maximum likelihood algorithms for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 8,* 128–141.

Erel, O., & Burman, B. (1995). Interrelatedness of marital relations and parent–child relations: A meta-analytic review. *Psychological Bulletin, 118,* 108–132.

Gonzales, N. A., Pitts, S. C., Hill, N. E., & Roosa, M. W. (2000). A mediational model of the impact of interparental conflict on child adjustment in a multiethnic, low-income sample. *Journal of Family Psychology, 14,* 365–379.

Greenberg, M. T., Speltz, M. L., & DeKlyen, M. (1993). The role of attachment in the early development of disruptive behavior problems. *Development and Psychopathology, 5,* 191–213.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis.* Cambridge, UK: Cambridge University Press.

Harter, S. (1983). Developmental perspectives on the self-system. In P. H. Mussen (Series Ed.) & E. M.

Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (4th ed., pp. 275–385). New York: Wiley.

Jacobvitz, D., Hazen, N., Curran, M., & Hitchens, K. (2004). Observations of early triadic family interactions: Boundary disturbances in the family predict symptoms of depression, anxiety, and attention-deficit/hyperactivity disorder in middle childhood. *Development and Psychopathology, 16*, 577–592.

Johnson, V. K. (2003). Linking changes in whole family functioning and children's externalizing behavior across the elementary school years. *Journal of Family Psychology, 17*, 499–509.

Johnson, V. K., Cowan, P. A., & Cowan, C. P. (1999). Children's classroom behavior: The unique contribution of family organization. *Journal of Family Psychology, 13*, 355–371.

Kerig, P. (1995). Triangles in the family circle: Effects of family structure on marriage, parenting, and child adjustment. *Journal of Family Psychology, 9*, 28–43.

Kretchmar, M. D., & Jacobvitz, D. B. (2002). Observing mother-child relationships across generations: Boundary patterns, attachment, and the transmission of caregiving. *Family Process, 41*, 351–374.

Ladd, G. W. (2003). Probing the adaptive significance of children's behavior and relationships in the school context: A child by environment perspective. *Advances in Child Behavior and Development, 31*, 43–104.

Ladd, G. W., & Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development, 72*, 1579–1601.

Ladd, G. W., & Profilet, S. M. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008–1024.

Lindahl, K. M., & Malik, N. M. (1996). *System for Coding Interactions and Family Functioning (SCIFF)*. Unpublished manual, University of Miami, Coral Gables, FL.

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment prediction tests: Their reliability and validity. *Marriage and Family Living, 21*, 251–255.

Lorber, M. F., & O'Leary, S. G. (2005). Mediated paths to overreactive discipline: Mothers' experienced emotion, appraisals, and physiological responses. *Journal of Consulting and Clinical Psychology, 73*, 972–981.

Malik, N. M., & Lindahl, K. M. (1996). *System for coding interactions in dyads (SCID)*. Unpublished manual, University of Miami, Coral Gables, FL.

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., et al. (2005). Developmental cascades: Linking academic achievement, externalizing and internalizing symptoms over 20 years. *Developmental Psychology, 41*, 733–746.

McHale, J., Johnson, D., & Sinclair, R. (1999). Family-level dynamics, preschoolers' family representations, and playground adjustment. *Early Education and Development, 10*, 373–401.

McHale, J. P., Kazali, C., Rotman, T., Talbot, J., Carleton, M., & Lieberson, R. (2004). The transition to coparenthood: Parents' prebirth expectations and early coparental adjustment at 3 months postpartum. *Development and Psychopathology, 16*, 711–733.

McHale, J. P., Kuersten-Hogan, R., & Lauretti, A. (1996). New directions in the study of family-level dynamics during infancy and early childhood. In J. P. McHale & P. A. Cowan (Eds.), *Understanding how family-level dynamics affect children's development: Studies of two-parent families* (pp. 5–26). San Francisco: Jossey-Bass.

McHale, J., Kuersten-Hogan, R., & Lauretti, A. (2001). Evaluating coparenting and family level dynamics during infancy and early childhood: The coparenting and family rating system. In P. Kerig & K. Lindahl (Eds.), *Family observational coding systems.* (pp. 151–170), Mahwah, NJ: Erlbaum.

McHale, J. P., Kuersten-Hogan, R., & Rao, N. (2004). Growing points for coparenting theory and research. *Journal of Adult Development, 11*, 221–234.

McHale, J., & Rasmussen, J. (1998). Coparental and family group-level dynamics during infancy: Early family precursors of child and family functioning during preschool. *Development and Psychopathology, 10*, 39–58.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Melby, J. N., & Conger, R. D. (2001). The Iowa Family Interaction Rating Scales: Instrument summary. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 33–58). Philadelphia: Brunner/Mazel.

Minuchin, S. (1974). *Families and family therapy*. Cambridge, MA: Harvard University Press.

Minuchin, P. (1985). Families and individual development: Provocations from the field of family therapy. *Child Development, 56*, 289–302.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.

Muthén, L., & Muthén, B. O. (2006). *Mplus* (4th ed.). Los Angeles: Author.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nylund, K. L., Asparouhov, T., & Muthen, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.

O'Connor, T. G., Hetherington, E. M., & Reiss, D. (1998). Family systems and adolescent development: Shared and nonshared risk and protective factors in nondivorced and remarried families. *Development and Psychopathology, 10*, 353–375.

Perry, K. E., & Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist*, *33*, 177–194.

Richmond, M. K., & Stocker, C. M. (2006). Associations between family cohesion and adolescent siblings' externalizing behavior. *Journal of Family Psychology*, *20*, 663–669.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Sroufe, L. A., Jacobvitz, D., Mangelsdorf, S., DeAngelo, E., & Ward, M. (1985). Generational boundary dissolution between mothers and their preschool children: A relationship systems approach. *Child Development*, *56*, 317–325.

Stoneman, Z., Brody, G. H., & Burke, M. (1989). Marital quality, depression and inconsistent parenting: Relationship with observed mother-child conflict. *American Journal of Orthopsychiatry*, *59*, 105–117.

Sturge-Apple, M. L., Davies, P. T., & Cummings, E. M. (2006). The impact of interparental hostility and withdrawal on parental emotional unavailability and children's adjustment difficulties. *Child Development*, *77*, 1623–1641.

Vecchiotti, S. (2003). Kindergarten: An overlooked educational policy priority. *Social Policy Report*, *17*, 3–17.

Webster-Stratton, C., & Hammond, M. (1999). Marital conflict management skills, parenting style, and early-onset conduct problems: Processes and pathways. *Journal of Child Psychology and Psychiatry*, *40*, 917–927.

Child Development, July/August 2010, Volume 81, Number 4, Pages 1336–1340

Manuscripts Accepted for Publication

Home Media and Children's Achievement and Behavior, Sandra Hofferth.

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance, Dante Cicchetti, Fred A. Rogosch, Mark L. Howe, and Sheree L. Toth.

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents, Tiffany Yip, Eleanor K. Seaton, and Robert M. Sellers.

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood, Elizabeth Votruba-Drzal, Rebekah Coley, Carolina Maldonado Carreno, Christine Li-Grining, and P. Lindsay Chase-Lansdale.

An Ecological Approach to Promoting Early Adolescent Mental Health and Social Adaptation: Family-Centered Intervention in Public Middle Schools, Elizabeth Stormshak, Arin M. Connell, Marie-Helene Veronneau, Michael W. Myers, Thomas J. Dishion, Kathryn Kavanagh, and Allison S. Caruthers.

Assessing the Effects of a Work-Based Antipoverty Program for Parents on Youth's Future Orientation and Employment Experiences, Rachel Kaplan, Vonnie C. McLoyd, Kelly Purtell, and Aletha Huston.

Children's Classroom Engagement and School Readiness Gains in Pre-Kindergarten, Nina C. Chien, Carollee Howes, Margaret Burchinal, Robert Pianta, Sharon Ritchie, Donna Bryant, Richard Clifford, Diane Early, and Oscar Barbarin.

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement, Katherine Pears, Philip A. Fisher, Jacqueline Bruce, Hyoun K. Kim, and Karen Yoerger.

Children's Understanding of Ordinary and Extraordinary Minds, Jonathan Lane, Henry Wellman, and E. Margaret Evans.

What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning, Christopher Fennell and Sandra R. Waxman.

Contagious Yawning in Autistic and Typical Development, Molly Helt, Inge-Marie Eigsti, Peter J. Snyder, and Deborah A. Fein.

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective, Maike Gieling, Jochem Thijs, and Maykel Verkuyten.

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning, Deborah L. Linebarger and Jessica Taylor Piotrowski.

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task, Sarah Berger, Karen E. Adolph, and Alisan E. Kavookjian.

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues, Kerri Johnson, Leah Lurye Murphy, and Louis G. Tassinary.

Gender Differences in Spatial Ability of Young Children: The Effects of Training and Processing Strategies, David Tzuriel and Gila Egozi.

Do Adolescents and Parents Reconstruct Memories About Their Conflict as a Function of Adolescent Attachment? Matthew Dykas, Susan S. Woodhouse, Katherine B. Ehrlich, and Jude Cassidy.

Cognitive Flexibility in Drawings of Bilingual Children, Esther Adi-Japha, Jennie Berberich-Artzi, and Afaf Libnawi.

The Contributions of Numerosity and Domain-General Abilities to School Readiness, Lynn Fuchs, David Geary, Donald Compton, Douglas Fuchs, Carol Hamlett, and Joan Bryant.

Prospective Associations Between Peer Victimization and Aggression, Jamie Ostrov.

Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning, Annie Bernier, Stephanie M. Carlson, Stéphanie Bordeleau, and Julie Carrier.

Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children, Stacey Doan and Qi Wang.

Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading, Serena Lecce, Silvia Zocchi, Adriano Pagnin, Paola Palladino, and Mele Taumoepeau.

Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups, Daphna Bassok.

Adolescents' Development of Skills for Agency in Youth Programs: Learning to Think Strategically, Reed Larson and Rachel Angus.

Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons, Olga Kochukhova and Gustaf Gredebäck.

Intergenerational Continuity in Child Maltreatment: Mediating Mechanisms and Implications for Prevention, Lisa Berlin, Karen Appleyard, and Kenneth Dodge.

Race Salience and Essentialist Thinking in Racial Stereotype Development, Kristin Pauker, Nalini Ambady, and Evan Apfelbaum.

Differing Levels of Gender Salience in Preschool Classrooms: Effects on Children's Gender Attitudes and Intergroup Bias, Lacey J. Hilliard and Lynn S. Liben.

Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children, Geraldine Legendre, Isabelle Barriere, Louise Goyet, and Thierry Nazzi.

Two-Year-Olds Use the Generic/Non-Generic Distinction to Guide Their Inferences About Novel Kinds, Susan Graham, Samantha Nayer, and Susan Gelman.

Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children, Stuart Marcovitch, Janet J. Boseovski, Robin J. Knapp, and Michael J. Kane.

The Social Costs of Academic Success Across Ethnic Groups, Thomas E. Fuller-Rowell and Stacey N. Doan.

Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Edentity, Thomas Lyon, Elizabeth A. Ahern, Lindsay A. Malloy, and Jodi A. Quas.

Understanding Bullying and Victimization During Childhood and Adolescence: A Mixed Methods Study, Nancy Guerra, Kirk R. Williams, and Shelly Sadek.

A Developmental Perspective on Executive Function, John R. Best and Patricia Miller.

Pathways to Mathematics: Longitudinal Predictors of Performance, Jo-Anne LeFevre, Lisa Fast, Sheri-Lynn Skwarchuk, Brenda Smith-Chant, Jeffrey Bisanz, Deepthi Kamawar, and Marcie Penner-Wilger.

Development in Children's Comprehension of Linguistic Register, Laura Wagner, Maia Greene-Havas, and Rebecca Gillespie.

How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition, Clarissa Thompson and John Opfer.

Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping, Margarita Svetlova, Sara R. Nichols, and Celia A. Brownell.

A Large-Scale Evaluation of the KiVa Anti-Bullying Program, Antti Kärnä, Marinus Voeten, Elisa Poskiparta, Todd D. Little, Anne Kaljonen, and Christina Salmivalli.

Movement Planning Reflects Skill Level and Age Changes in Toddlers, Yu-ping Chen, Rachel Keen, Kerstin Rosander, and Claes von Hofsten.

Young Children Selectively Avoid Helping People With Harmful Intentions, Amrisha Vaish, Malinda Carpenter, and Michael Tomasello.

Danish Majority Children's Reasoning About Exclusion Based on Gender and Ethnicity, Signe Juhl Moller and Harriet Tenenbaum.

How Does It Look? Level 2 Perspective-Taking at 36 Months of Age, Henrike Moll and Andrew Meltzoff.

Emotion Discourse, Social Cognition, and Social Skills in Children With and Without Developmental Delays, Rachel M. Fenning, Bruce L. Baker, and Jaana Juvonen.

A Monozygotic Twin Differences Study of Friends' Aggression and Children's Adjustment Problems, Frank Vitaro, Mara Brendgen, Michel

Boivin, Stephane Cantin, Ginette Dionne, Richard E. Tremblay, Alain Girard, and Daniel Perusse.

Gender Differences in Adolescents' Daily Interpersonal Events and Well-Being, Lisa Flook.

Teacher–Child Relationships and the Development of Academic and Behavioral Skills During Elementary School: A Within- and Between-Child Analysis, Carolina Maldonado-Carreño and Elizabeth Votruba-Drzal.

The Effects of the Fast Track Preventive Intervention on the Development of Conduct Disorder Across Childhood, Conduct Problems Prevention Research Group.

Object Similarity Bootstraps Young Children to Action-Based Verb Extensions, Etsuko Haryu, Mutsumi Imai, and Hiroyuki Okada.

The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions, Joseph Durlak, Roger Weissberg, Allison Dymnicki, Rebecca Taylor, and Kriston Schellinger.

Science Does Not Speak for Itself: Translating Child Development Research for the Public and Its Policy Makers, Jack Shonkoff and Susan Nall Bales.

Domain General Constraints on Statistical Learning, Erik Thiessen.

Reciprocal Relations Between Parenting and Adjustment in a Sample of Juvenile Offenders, Lela Rankin Williams and Laurence Steinberg.

A Randomized Controlled Trial of Child FIRST: A Comprehensive Home-Based Intervention Translating Research Into Early Childhood Practice, Darcy I. Lowell, Alice S. Carter, Leandra Godoy, Belinda Paulicin, and Margaret J. Briggs-Gowan.

Paths of Effects From Preschool to Adult Well-Being: A Confirmatory Analysis of the Child–Parent Center Program, Arthur Reynolds and Suh-Ruu Ou.

Age 26 Cost–Benefit Analysis of the Child–Parent Center Early Education Program, Arthur Reynolds, Judy Temple, Barry White, Suh-Ruu Ou, and Dylan Robertson.

Latino Adolescents' Experiences of Discrimination Across the First Two Years of High School: Correlates and Influences on Educational Outcomes, Aprile Benner and Sandra Graham.

The Generic/Non-Generic Distinction Influences How Children Interpret New Information About Social Others, Andrei Cimpian and Ellen M. Markman.

Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research, Stephanie Jones, Joshua Brown, and J. Lawrence Aber.

Conditional Reasoning With False Premises Facilitates the Transition Between Familiar and Abstract Reasoning, Henry Markovits and Hugues Lortie Forgues.

When Easy Comes Hard: The Development of Adaptive Strategy Selection, Rui Mata, Bettina von Helversen, and Joerg Rieskmap.

Coherent Motion Sensitivity and Reading Development in the Transition from Pre-Reading to Reading Stage, Bart Boets, Maaike Vandermosten, Piers Cornelissen, Jan Wouters, and Pol Ghesquière.

Same-Ethnic Friendship Preferences Among German and Turkish Preadolescents Attending Ethnically Heterogeneous Schools, Philipp Jugert, Peter Noack, and Adam Rutland.

Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect, Chiara Turati, Elisa Di Giorgio, Lara Bardi, and Francesca Simion.

Math-Gender Stereotypes in Elementary School Children, Dario Cvencek, Andrew N. Meltzoff, and Anthony G. Greenwald.

Reporter Discrepancies Among Parents, Adolescents, and Peers: Adolescent Attachment and Informant Depressive Symptoms as Explanatory Factors, Katherine B. Ehrlich, Jude Cassidy, and Matthew J. Dykas.

Urban and Rural Chinese Adolescents' Judgments and Reasoning About Personal and Group Jurisdiction, Charles Helwig, Shaogang Yang, Dingliang Tan, Chunqiong Liu, and Tiffany Shao.

Childhood Trauma and Psychiatric Disorders as Correlates of School Dropout in a National Sample of Young Adults, Michelle V. Porche, Lisa R. Fortuna, Julia Lin, and Margarita Alegria.

The Role of Peer Relationships in the Development of Early School-Age Externalizing Problems, Cristina Sturaro, Pol A. C. van Lier, Pim Cuijpers, and Hans M. Koot.

Relationship Between Language Experiences in Preschool Classrooms and Children's Kindergarten and Fourth-Grade Language and Reading Abilities, David K. Dickinson and Michelle V. Porche.

Intergenerational Transmission of Adaptive Functioning: A Test of the Interactionist Model of SES and Human Development, Thomas Schofield, Monica Martin, Katherine Conger, Brent Donnellan, Tricia Neppl, and Rand Conger.

Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring, Carla Herrera, Jean Baldwin Grossman, Tina J. Kauh, and Jennifer McMaken.

Remission of Depression in Parents: Links to Healthy Functioning in Their Children, Judy Garber, Jeffrey A. Ciesla, Elizabeth McCauley, Guy Diamond, and Kelly A. Schloredt.

Double Jeopardy: Poorer Social-Emotional Outcomes for Children in the NICHD SECCYD Experiencing Home and Child Care Environments That Confer Risk, Sarah Watamura, Deborah Phillips, Taryn Morrissey, Kathleen McCartney, and Kristen Bub.

Imagining a Way Out of the Gravity Bias: Preschoolers Can Visualize the Solution to a Spatial Problem, Amy Joh, Vikram Jaswal, and Rachel Keen.

Sequential Progressions in a Theory of Mind Scale: Longitudinal Perspectives, Henry Wellman, Fuxi Fang, and Candida Peterson.

Can Maltreated Children Inhibit True and False Memories for Emotional Information? Mark Howe, Sheree Toth, and Dante Cicchetti.

Parents' Involvement in Children's Learning in the United States and China: Implications for Children's Academic and Emotional Adjustment, Cecilia Cheung and Eva Pomerantz.

Four Children and One Toy: Chinese and Canadian Children Faced With Potential Conflict Over a Limited Resource, Doran French, Xinin Chen, Janet Chung, Miao Li, Dan Li, Huichang Chen, and Li Wang.

Influence of Eye Gaze on Spoken Word Processing: An ERP Study With Infants, Eugenio Parise, Andrea Handl, Letizia Palumbo, and Angela Friederici.

Infant and Maternal Sensitivity to Interpersonal Timing, Anne Henning and Tricia Striano.

Emotion Regulation Profiles, Temperament, and Adjustment Problems in Preadolescents, Maureen Zalewski, Liliana J. Lengua, Anna C. Wilson, Anika Trancik, and Alissa Bazinet.

Developmental Pathways to Integrated Social Skills: The Roles of Parenting and Early Intervention in Raising Healthy Children, Catherine Ayoub, Claire Vallotton, and Ann Mastergeorge.

Consequences of ''Minimal'' Group Affiliations in Children, Yarrow Dunham, Andrew Baron, and Susan Carey.

Revisiting the Impact of Part-Time Work on Adolescent Adjustment: Distinguishing Between Selection, Kathryn C. Monahan, Joanna Lee, and Laurence Steinberg.

Parents as Moderators of the Impact of School Norms and Peer Influences on Aggression in Middle School Students, Albert Farrell, David B. Henry, Sally Mays, and Michael E. Schoeny.

The ABCs of Family Mealtimes: Observational Lessons for Promoting Healthy Outcomes for Children With Persistent Asthma, Barbara Fiese, Marcia A. Winter, and Joanna Corra Botti.

Language Development in Internationally Adopted Children: A Special Case of Early Second Language Learning, Karine Gauthier and Fred Genesee.

Third-Party Social Interaction and Word Learning From Video, Kate O'Doherty, Georgene L. Troseth, Priya Shimpi, Elizabeth Goldenberg, Nameera Akhtar, and Megan M. Saylor.

Computer-Assisted Remedial Reading Intervention for School Beginners at Risk for Reading Disability, Nina Saine, Marja-Kristiina Lerkkanen, Timo Ahonen, Asko Tolvanen, and Heikki Lyytinen.

Desire for Increased Autonomy and Adolescents' Perceptions of Peer Autonomy: ''Everyone Else Can; Why Can't I?'' Christopher Daddis.

Almost Thinking Counterfactually: Children's Understanding of Close Counterfactuals, Sarah Beck and Carlie Guthrie.

CSRP's Impact on Low-Income Preschoolers' Pre-Academic Skills: Self-Regulation as a Mediating Mechanism, C. Cybele Raver, Stephanie

M. Jones, Christine Li-Grining, Fuhua Zhai, Kristen Bub, and Emily Pressler.

Does God Make It Real? Children's Belief in Religious Stories From the Judeo-Christian Tradition, Victoria Cox Vaden and Jacqueline D. Woolley.

Changes in Early Adolescents' Sense of Responsibility to Their Parents in the United States and China: Implications for Their Academic Functioning, Eva Pomerantz, Lili Qin, Qian Wang, and Huichang Chen.

Two- to Eight-Month-Old Infants' Perception of Dynamic Auditory-Visual Spatial Co-Location, Gavin Bremner, Alan Slater, Scott Johnson, Uschi Mason, Jo Spring, and Maggie Bremner.

Longitudinal Associations Between Marital Instability and Child Sleep Problems Across Infancy and Toddlerhood in Adoptive Families, Anne Mannering, Gordon Harold, Leslie Leve, Katherine Shelton, Daniel Shaw, Rand Conger, Jenae Neiderhiser, Laura Scaramella, and David Reiss.

Promoting Effective Parenting Practices and Preventing Child Behavior Problems in School, Laurie Brotman, Esther Calzada, Keng-Yen Huang, Sharon Kingston, Spring Dawson-McClure, Dimitra Kamboukos, Amanda Rosenfelt, Ami Schwab, and Eva Petkova.

Infantile Amnesia Across the Years: A 2-Year Follow-Up of Children's Earliest Memories, Carole Peterson, Kelly L. Warren, and Megan M. Short.

Hemoglobin, Growth, and Attention of Infants in Southern Ethiopia, Nicki Aubuchon-Endsley, Stephanie L. Grant, Getenesh Berhanu, David G. Thomas, Sarah Schrader, Devon Eldridge, Tay S. Kennedy, and Michael Hambidge.

Transition From Crawling to Walking and Infants' Actions With Objects and People, Lana Karasik, Catherine Tamis-LeMonda, and Karen Adolph.

The Genetic and Environmental Covariation Among Psychopathic Personality Traits and Reactive and Proactive Aggression in Childhood, Serena Bezdjian, Catherine Tuvblad, Adrian Raine, and Laura A. Baker.

fellowships available

The Radcliffe Institute for Advanced Study at Harvard University awards 40 funded residential fellowships each year designed to support scholars, scientists, artists, and writers of exceptional promise and demonstrated accomplishment.

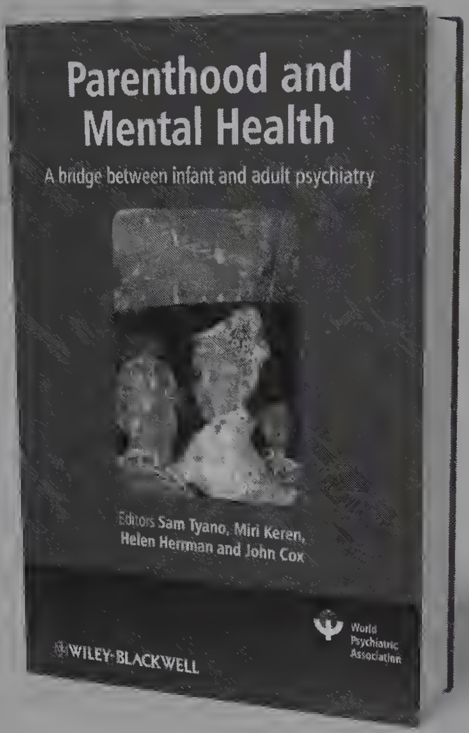
For more information, please contact:
Radcliffe Application Office
8 Garden Street
Cambridge, MA 02138
617-496-1324
fellowships@radcliffe.edu
www.radcliffe.edu

RADCLIFFE INSTITUTE FOR ADVANCED STUDY
HARVARD UNIVERSITY

Parenthood and Mental Health
A bridge between infant and adult psychiatry

EDITED BY

Sam Tyano, Tel Aviv University, Sackler School of Medicine, Israel
Miri Keren, Geha Mental Health Center, Petach Tikva, Israel
Helen Herrman, University of Melbourne, Australia
John Cox, University of Gloucestershire, UK


Parenthood and Mental Health
A bridge between infant and adult psychiatry
Editors Sam Tyano, Miri Keren, Helen Herrman and John Cox
WILEY-BLACKWELL
World Psychiatric Association

Across all cultures parenting is the foundation of family life. It is the domain where adult mental health meets infant development. Beginning in pregnancy, parenting involves many conscious and unconscious processes which have recently been shown to affect a child's development significantly. This book focuses on pregnancy and the first year of life, providing a thorough account of the points of encounter between adult and infant psychiatry. In a fresh and comprehensive way, it summarises knowledge about early parenting, including a critical analysis of parenting, what it means to be a "good enough parent", and its relationship to infant, parent and family outcomes. In addition to the psychiatric dimension, the book emphasises the biological aspects of parenting, parental psychopathology and normal and abnormal infant development.

Hardback: 978-0-470-74722-3
April 2010 • 484 pages

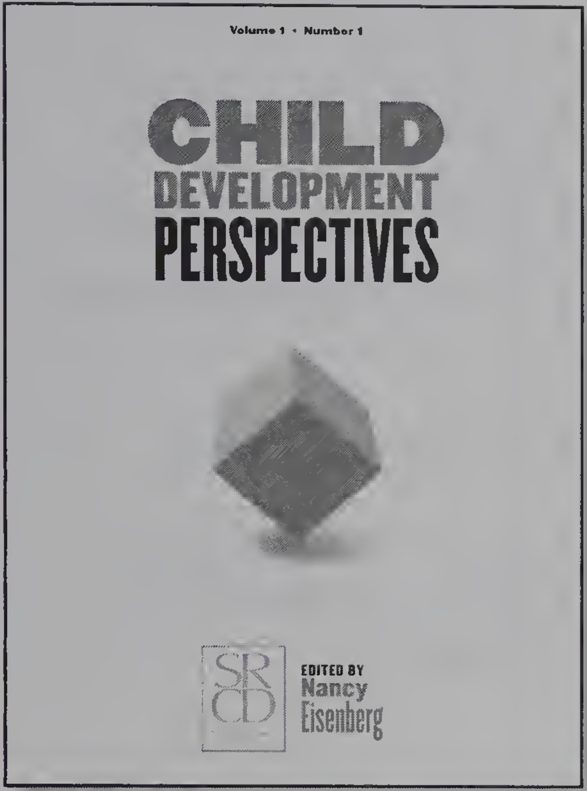
E-book: 978-0-470-66067-6

"Tyano, Keren, Herrman and Cox have edited a thoughtfully prepared guide on normal and abnormal parenting. They have, with enormous skill and wisdom, helped to unite the important aspects of pregnancy, infant and childhood development and parenting for adult and child and adolescent psychiatrists. World-class internationally recognized clinicians and researchers help make this book useful throughout the world. This is a masterful, culturally sensitive and important book which provides a long overdue and much needed guide on relationships among children, parents and families."

MICHELLE RIBA
Professor and Associate Chair for Integrated Medical and Psychiatric Services
Department of Psychiatry, University of Michigan, USA

WILEY-BLACKWELL

For more information or to place an order visit
www.wiley.com/go/psychiatry


Volume 1 · Number 1

CHILD DEVELOPMENT PERSPECTIVES

SRCD

EDITED BY Nancy Eisenberg

CHILD
DEVELOPMENT
PERSPECTIVES

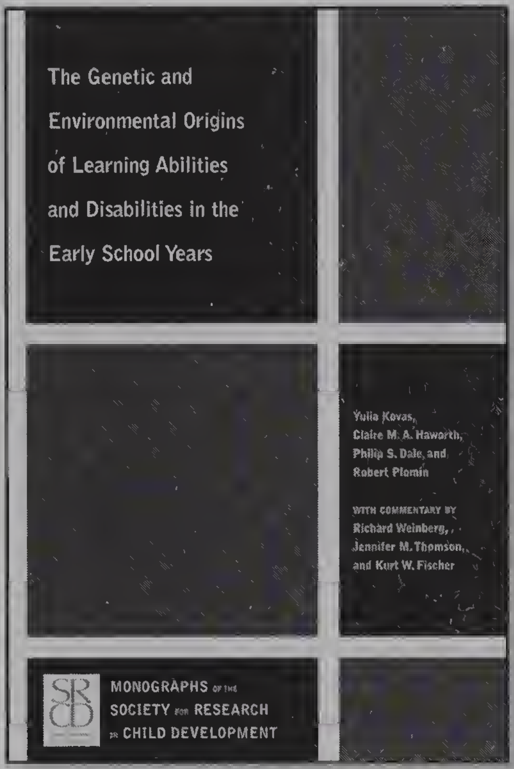
Edited by
NANCY EISENBERG

Published on behalf of the Society for Research in Child Development

Child Development Perspectives' mission is to provide accessible, synthetic reports that summarize emerging trends or conclusions within various domains of developmental research, and to encourage multidisciplinary and international dialogue on a variety of topics in the developmental sciences.

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.interscience.wiley.com/journal/cdep



The Genetic and
Environmental Origins
of Learning Abilities
and Disabilities in the
Early School Years

Yulia Kovas,
Claire M. A. Haworth,
Philip S. Dale, and
Robert Plomin

WITH COMMENTARY BY
Richard Weinberg,
Jennifer M. Thomson,
and Kurt W. Fischer

SRCD
MONOGRAPHS OF THE
SOCIETY FOR RESEARCH
IN CHILD DEVELOPMENT

MONOGRAPHS OF THE
SOCIETY FOR RESEARCH
IN CHILD DEVELOPMENT

Edited by
W. ANDREW COLLINS

Published on behalf of the Society for Research in Child Development

Since 1935 this series has presented in-depth research studies and significant findings in child development and its related disciplines. Each issue consists of a single study or a group of papers on a single theme, accompanied usually by commentary and discussion. Like all Society for Research in Child Development (SRCD) publications, the ***Monographs*** enable development specialists from many disciplines to share their data, techniques, research methods, and conclusions.

A subscription to the ***Monographs*** series also includes a full subscription (6 issues) to ***Child Development,*** the flagship journal of the SRCD, and ***Child Development Perspectives,*** the most recent publication from SRCD.

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.interscience.wiley.com/journal/mono

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION
- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial No. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY
- *Mona El-Sheikh et al.* (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN
- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING
- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE
- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS
- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE
- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT
- *Michelle M. Chouinard* (Serial No. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS
- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM
- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997
- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD
- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS
- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT
- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES
- *Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES
- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR
- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

ORIGINS AND EARLY DEVELOPMENT OF HUMAN BODY KNOWLEDGE
- *Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

MOTHER CHILD CONVERSATIONS ABOUT GENDER: UNDERSTANDING THE ACQUISITION OF ESSENTIALIST BELIEFS
- *Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

WILEY-BLACKWELL

SEPTEMBER/OCTOBER 2010 • VOLUME 81, NUMBER 5

CHILD DEVELOPMENT

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

SRCD
Society for Research
in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit wileyonlinelibrary.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley Online Library. Visit wileyonlinelibrary.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to wileyonlinelibrary.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at http://wileyonlinelibrary.com/support
Visit our Online Customer Self-Help available in 6 languages at http://onlinelibrary.custhelp.com

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Child Development © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor
Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor
Detra N. Davis

Editorial Assistant
Adam Martin

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Ramesh Bhatt
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Cara Cashon
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Anna Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Greg Reynolds
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, September/October 2010, Volume 81, Number 5, Pages 1341–1636

Contents

IN THIS ISSUE 1341

EMPIRICAL REPORTS

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues
Kerri L. Johnson, Leah E. Lurye, and Louis G. Tassinary 1346

Cognitive Flexibility in Drawings of Bilingual Children
Esther Adi-Japha, Jennie Berberich-Artzi, and Afaf Libnawi 1356

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task
Sarah E. Berger, Karen E. Adolph, and Alison E. Kavookjian 1367

What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning
Christopher T. Fennell and Sandra R. Waxman 1376

EMPIRICAL ARTICLES

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective
Maike Gieling, Jochem Thijs, and Maykel Verkuyten 1384

The Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study
Elizabeth Pellicano 1400

Gender Differences in Spatial Ability of Young Children: The Effects of Training and Processing Strategies
David Tzuriel and Gila Egozi 1417

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents
Tiffany Yip, Eleanor K. Seaton, and Robert M. Sellers 1431

Do Adolescents and Parents Reconstruct Memories About Their Conflict as a Function of Adolescent Attachment?
Matthew J. Dykas, Susan S. Woodhouse, Katherine B. Ehrlich, and Jude Cassidy 1445

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood
Elizabeth Votruba-Drzal, Rebekah Levine Coley, Carolina Maldonado-Carreño, Christine P. Li-Grining, and P. Lindsay Chase-Lansdale 1460

Children's Understanding of Ordinary and Extraordinary Minds
Jonathan D. Lane, Henry M. Wellman, and E. Margaret Evans 1475

Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children
Stacey N. Doan and Qi Wang 1490

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance
Dante Cicchetti, Fred A. Rogosch, Mark L. Howe, and Sheree L. Toth 1504

The Contributions of Numerosity and Domain-General Abilities to School Readiness
Lynn S. Fuchs, David C. Geary, Donald L. Compton, Douglas Fuchs, Carol L. Hamlett, and Joan D. Bryant 1520

Children's Classroom Engagement and School Readiness Gains in Prekindergarten
Nina C. Chien, Carollee Howes, Margaret Burchinal, Robert C. Pianta, Sharon Ritchie, Donna M. Bryant, Richard M. Clifford, Diane M. Early, and Oscar A. Barbarin 1534

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement
Katherine C. Pears, Philip A. Fisher, Jacqueline Bruce, Hyoun K. Kim, and Karen Yoerger 1550

On the Progression and Stability of Adolescent Identity Formation: A Five-Wave Longitudinal Study in Early-to-Middle and Middle-to-Late Adolescence
Wim Meeus, Rens vad de Schoot, Loes Keijsers, Seth J. Schwartz, and Susan Branje 1565

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning
Deborah L. Linebarger and Jessica Taylor Piotrowski 1582

Home Media and Children's Achievement and Behavior
Sandra L. Hofferth 1598

Contagious Yawning in Autistic and Typical Development
Molly S. Helt, Inge-Marie Eigsti, Peter J. Snyder, and Deborah A. Fein 1620

OTHER

Manuscripts Accepted 1632

Child Development, September/October 2010, Volume 81, Number 5, Pages 1341–1345

In This Issue

Eye on Gender

From early in life, people categorize others according to the social group they belong to, for example, by gender, race, or age. For preschoolers, these determinations often involve gender, with their judgments influenced by superficial aspects of appearance like hairstyle and dress. **Johnson, Lurye, and Tassinary** (p. 1346) find that 4-year-olds' judgments about whether the shape they viewed was male or female were unrelated to body shape. But by age 5, the youngsters had started to rely on more permanent aspects of appearance such as body shape to categorize. And by age 6, their judgments resembled those of adults. Further analyses of how children looked at the shapes find that when they looked toward the waist and hip region, they used that information to make their decisions. The study, which used eye-tracking technology to determine where the more than 100 children studied were looking, furthers our understanding of how youngsters view others.

Bilingual Flexibility

Bilingual children speak two languages and can swiftly move from the use of one language to another. Research has shown that these children develop improved abilities that give them flexibility in learning and using both languages. **Adi-Japha, Berberich-Artzi, and Libnawi** (p. 1356), in a study of 4-year-old English-Hebrew children and 5-year-old Arabic-Hebrew children, find that these advantages in cognitive flexibility even extend to areas beyond language, such as imaginative drawings, with bilingual children showing a form of expressive drawing earlier than monolingual children. The study sheds light on the relations between processes involved in language learning and other symbolic domains.

Spatial Ability

We use spatial ability in our daily lives when we navigate, recall where something is located, and recognize things, and in jobs like engineering, architecture, and surgery. Past research has found that males have better spatial ability than females, with the gender gap starting in infancy and growing with age. In a study of more than 100 first graders in Israel, **Tzuriel and Egozi** (p. 1417) asked: Can training reduce these gender differences? They find that after a relatively long-term training program (the study comprised eight weekly sessions) focused on expanding working memory, perceiving spatial information globally rather than analytically, and thinking about spatial stimuli from different points of view, initial gender differences disappeared. And later, girls who had taken part in the training did better than boys who hadn't. Given the value of good spatial skills in math and science, this study can inform curriculum and intervention programs aimed at overcoming gender differences that many believe are biologically determined.

Researchers typically look at how babies solve problems by examining how they use manual tools and considering spatial relations between the baby, the tool, and the goal at hand. **Berger, Adolph, and Kavookjian** (p. 1367) took a different approach, focusing on the whole bodies of the 16-month-olds studied, not just their hands. They find that youngsters were also able to solve tasks that involved navigating bridges and handrails, suggesting that they could plan solutions that required them to use their whole bodies and to coordinate multiple motor actions in the face of ongoing challenges. The study extends our understanding of early cognitive abilities in young children.

Linguistics: Putting Words in Context

When babies first learn words, do they notice subtle sound differences? Can they distinguish words like *bad* and *dad* when learning their meanings? In the past, evidence suggested that 14-month-olds don't bring their detailed attention to language sounds to bear in learning words. But **Fennell and Waxman** (p. 1376) changed the way the words were presented, making clear that the words were nouns and presenting them in a naming context. They find

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0001

that bringing in the notion of reference—a basic feature of human language—changes our understanding of how infants learn words. The findings tell us that infants clearly use sounds that are relevant to the language they're learning (such as the consonants *b* and *d* in English) and that they refine those pertinent sounds over the first year of life to help them learn words. The study can inform language-teaching strategies.

Teens: Racial Identity and Cultural Diversity

In a longitudinal study, **Yip, Seaton, and Sellers** (p. 1431) asked more than 200 Black teens in predominantly White middle and high schools in the Midwestern United States to complete surveys about their friendships and racial identity. They find that Black youths' racial identity—that is, the significance of race in their everyday lives—was more likely to change over the 3 years of the study if they had close Black friends. Racial identity remained stable for Black youths who didn't have a lot of contact with other Black youths. The research suggests that school makeup interacts with friendships to affect adolescents' racial identity, and that having friends of the same race is important for the development of racial identity. Since racial identity has been shown to be a protective factor for mental health, this study has implications for educators.

In the Netherlands, prejudice toward Muslims is said to be relatively high, with half of the participants in a recent nationwide survey saying Muslims threaten the Dutch way of life. In a study with global significance, **Gieling, Thijs, and Verkuyten** (p. 1384) looked at more than 1,000 ethnically Dutch non-Muslim teens to gauge their tolerance toward such current issues as Muslims wearing headscarves and founding separate Islamic schools. They also measured the extent to which the teens believed that cultural diversity is good for society and should be accommodated equitably. The teens expressed moderate to low levels of tolerance, with levels declining with age. While some think that as people get older, their reasoning grows more principled and they become more tolerant, this study finds just the opposite. This could be because older children are increasingly concerned about how Dutch society functions and tend to see Muslim practices as threatening. By shedding light on the different ways adolescents reason about these issues, the study has implications for intervention and education programs.

Insights Into Autism

People with autism are thought to have a specific profile of cognitive strengths and weaknesses, but few studies have tracked children's cognitive skills over time. In a longitudinal study of about 70 children, **Pellicano** (p. 1400) finds that compared with children who don't have autism, children with autism have difficulties appreciating the mental lives of others, problems regulating and controlling their behavior, as well as an enhanced ability to perceive details at age 5–6 and again 3 years later. Yet a closer look at individual children's patterns of performance reveals that not all children with autism display the same profile of cognitive strengths and weaknesses, indicating that profiles of cognitive skills vary from one child to the next. The study also finds that some of the cognitive skills of the children with autism improve over time. The findings help inform our understanding of autism.

When someone near you yawns, it's hard to keep yourself from yawning. Research backs this concept, noting that about half of typical adults yawn after someone else has done so. But what about children—do they do the same? In their study of 120 typically developing 1- to 6-year-olds, **Helt, Eigsti, Snyder, and Fein** (p. 1620) find that although infants begin to yawn spontaneously even before they leave the womb, most children don't appear to be susceptible to contagious yawning until they're about 4. The team also carried out the same study on about thirty 6- to 15-year-olds with autism spectrum disorder (ASD), comparing them to two other groups of children with the same mental and chronological ages. They find that children with ASD are less likely to yawn contagiously than typically developing children, and those children with diagnoses that imply more severe autistic symptoms are much less likely to yawn contagiously than those with milder diagnoses. Given that contagious yawning may signal empathy, this study suggests that empathy (and the mimicry that may underlie it) develops slowly over the first few years of life—and that children with ASD may miss subtle cues that tie them emotionally to others.

Parent–Child Interactions: Talking and Remembering

Past research has shown that the ways mothers talk to their children influence children's understanding of themselves and others. Specifically, mothers who focus on discussing desires, feelings, and thoughts contribute to the way children understand how psychological states like emotions, beliefs, and desires influence behaviors. **Doan and Wang** (p. 1490) asked: Can styles of conversation influence children's understanding of how different situations may elicit different types of emotions? Are there cultural differences, and can they explain differences in children's understanding of emotions? The researchers studied more than 130 middle-class European American and immigrant Chinese moms and their 3-year-olds. Findings: European American mothers were more likely to talk about desires, thoughts, and emotions and to ask their children to talk about their psychological states (such as their feelings, beliefs, and thoughts). Immigrant Chinese mothers, on the other hand, were more likely to talk about behaviors. European American children rated higher than their immigrant Chinese counterparts on tests that measured how well they understood the causes of certain emotions. The findings add to our understanding of the cultural and universal processes underlying emotional development.

How do teens and parents remember specific parts of a disagreement after it's over, is their memory accurate, and does it mesh with what really happened? Over time, as memories fade, how do they fill in the gaps? **Dykas, Woodhouse, Ehrlich, and Cassidy** (p. 1445) looked at almost 200 teens in 11th grade and their parents to determine whether the answers to these questions have anything to do with the kinds of attachment relationships the teens had with their parents. They find that secure teens (those with positive models of attachment) remembered interactions more favorably over time, while insecure teens (those with negative models) showed less favorable memory. A similar pattern held for moms, but not for dads, in relation to their teens' adjustment status. The findings suggest that memories are remembered over time based on how adolescents mentally represent previous experiences with their parents.

Child Care: Quality Matters

More than 12 million U.S. preschoolers are in child care. The findings of a longitudinal study by **Votruba-Drzal, Coley, Maldonado-Carreño, Li-Grining, and Chase-Lansdale** (p. 1460) suggest that child care may have a greater effect on some children than others. The researchers looked at about 350 children from low-income ethnic-minority families when they were preschoolers and again in middle childhood. Based on their child care settings, the study finds that children who attended more responsive, stimulating, and well-structured settings during preschool had fewer externalizing behavior problems (such as being aggressive and breaking rules) in middle childhood. High-quality child care was particularly important for boys and African American children. In highlighting the importance of high-quality child care for low-income children, the study strengthens calls for policy and programmatic efforts to increase low-income families' access to such care.

In many early childhood education programs, children are free to choose from a wide variety of educational materials to play with—this is called free-choice play. **Chien et al.** (p. 1534) studied more than 2,700 kindergarteners (almost 50% of whom were poor) in 11 U.S. states. They find that children who spend much of their classroom days engaged in free-choice play make smaller gains in language and math, compared with children who get either direct instruction from teachers or so-called scaffolding interactions (in which teachers modify their instruction to fit children's changing needs). Children who are given time to work on their own and are guided benefit the most, particularly poor children. The study has implications for policy and practice, suggesting that more quality instructional time spent with teachers, and less free play time without teacher guidance, prepares children better for starting school.

Superheroes, Deities, and More

Most people have ideas about nonhuman or superhuman beings that have extraordinary mental capacities, such as people who seem to have vast stores of knowledge, fictional super-heroes, or all-knowing deities. **Lane, Wellman, and Evans** (p. 1475) sought to determine how and when such concepts develop. They asked 3- to 5-year-olds questions to assess their understanding of extraordinary mental capacities in a variety of beings—including ordinary people and those with extraordinary capacities. They find that 4-year-olds first gain an understanding of human mental abilities (with all their attendant constraints) before they begin to appreciate the extraordinary mental abilities of nonhuman or super-human beings. In other words, children can't appreciate extraordinary mental abilities before they understand the limitations of ordinary minds. The research helps us further understand children's concepts of others, both ordinary and extraordinary.

Child Maltreatment

When they reach school, children in foster care run the risk of such problems as getting failing grades, behaving badly, and dropping out. However, not much is known about how and when these problems develop. **Pears, Fisher, Bruce, Kim, and Yoerger** (p. 1550) studied about 140 children aged 3–6 years—85 of whom were placed in foster care because of maltreatment; the rest lived with their biological parents and had no record of maltreatment. They find that for the foster kids, the problems developed early: By kindergarten and first grade, the foster children were markedly behind their peers in academics and behavior. But the study also finds that foster children with better self-control had better scores in both academics and on behavior. And children whose caregivers were more involved in school also had better scores on behavior. The findings can inform intervention programs to help maltreated foster children before they begin school.

Cicchetti, Rogosch, Howe, and Toth (p. 1504) looked at how children who have been maltreated remember. In a study of more than 300 low-income elementary schoolers, they sought to determine whether maltreated children had different levels of the stress hormone cortisol, and whether atypical levels of cortisol affected the children's basic memory abilities. They find that maltreated children were more likely to have low levels of cortisol in the morning (which can indicate overwhelming stress) but that the experience of child maltreatment per se didn't adversely affect their basic ability to remember. However, among children who'd been neglected or emotionally maltreated, those with low cortisol levels had more memory problems and weren't as accurate identifying words they had seen before. The findings suggest that maltreatment doesn't affect all children in the same way, and that maltreated children in general don't have significant problems with memory. By shedding light on these issues, the study has implications for both intervention programs and forensic investigations.

Boosting Numerical Abilities

Fuchs, Geary, Compton, Fuchs, Hamlett, and Bryant (p. 1520) assessed how basic numerical cognition and several general aspects of cognition, including working memory and reasoning, affect first graders' skills at solving arithmetic and word problems. Results suggest that children's arithmetic and word-problem skills depend on different groupings of numerical versus more general cognitive abilities and that providing opportunities to children to strengthen certain abilities can help build a foundation for formal mathematics learning in school. The study, of more than two hundred 5- to 7-year-olds, underscores the value of starting early—with preschoolers—to help children organize and strengthen their knowledge about numerical values.

Identity in Adolescence

A key component of adolescence is developing a sense of identity. But is identity a stable individual disposition, or does it change during the teenage years? **Meeus, van de Schoot, Keijsers, Schwartz, and Branje** (p. 1565), in a longitudinal study of the development of identity in more than 1,300 Dutch youths aged 12–20, find that identity progresses systematically during adolescence. Their study finds that most teens remain in the same identity over time, and most teens who change identity status over 5 years make only one transition, both of which indicate that changes in identity status tend to be decisive. The research suggests that adolescent identity develops at a regular pace and that, contrary to common assumptions, the majority of adolescents don't experience identity conflicts very often. On the other hand, a substantial group (about 13%) struggles with an insecure identity throughout adolescence; they are the youngsters at risk.

Media: New Views of Screen Time

Educational TV like *Sesame Street* can support children's academic and social-emotional development. **Linebarger and Piotrowski** (p. 1582) conducted two studies to identify features of educational programs that help children understand the programs and build vocabulary. They did a content analysis of learning strategies embedded in educational TV. Then they explored how more than 70 low-income African Americans and English language learners aged 7 through 9 did on tests of vocabulary and comprehension that were related to what they saw on the programs. Two types of shows were considered: narrative programs (those that tell stories to present information) and expository programs (those that present information in a variety of ways). The study concludes: Children learn vocabulary and understand program content better when they watch narrative programs than when they watch expository programs. The findings can help those creating educational TV programs designed to support at-risk youngsters' literacy skills.

Another study addresses a different type of screen time. Between 1997 and 2003, the amount of playtime children spent logged onto the computer and playing video games rose from 21% to 33%. Amid concerns about the effect of unlimited use of new media, **Hofferth** (p. 1598) asked: Does computer and video game use at home contribute to or detract from children's achievement and behavior? Using data from a nationally representative study of children aged 6–12 in 1997 who were followed for 6 years, she finds that the answers to this question differ depending on the children's race and ethnicity, and their gender. More computer time was associated with higher achievement for White and Black girls and Black boys, but not for White boys. Black girls who played more video games were better at solving math problems, but increased game playing meant lower verbal achievement on average for both groups of girls; there were no effects on boys' achievement. And for both groups of boys, more video playing was linked to more aggressive behavior. The study's conclusion: Although bringing minority boys and majority and minority girls up to speed on the computer could boost achievement, it's still important to monitor their video game time.

Anne Bridgman

Child Development, September/October 2010, Volume 81, Number 5, Pages 1346–1355

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues

Kerri L. Johnson
University of California

Leah E. Lurye
New York University

Louis G. Tassinary
Texas A&M University

Two studies examined how children between ages 4 and 6 use body shape (i.e., the waist-to-hip-ratio [WHR]) for sex categorization. In Study 1 ($N = 73$), 5- and 6-year-olds, but not 4-year-olds, selected bodies with increasingly discrepant WHRs to be "most like a man" and "most like a woman." Similarly, sex category judgments made by 5- and 6-year-olds, but not 4-year-olds, varied with WHR. In Study 2 ($N = 41$), eye movements indicated the functional use of waist and hips in sex categorization. Visual scanning behavior predicted the degree of association between WHR and judgment. Collectively, these results suggest that the ability to exploit sexual dimorphism to compel categorization develops between the ages of 4 and 6. Implications for theories of gender development and psychological essentialism are discussed.

Snips and snails, and puppy dogs' tails
That's what little boys are made of!. . .
Sugar and spice and all things nice
That's what little girls are made of!

In the popular rhyme, Mother Goose provided an endearing image about the makeup of little boys and girls. The rhyme's appeal rests, in part, on its vivid analogies for the distinction between boys and girls. Although the content is indisputably false, two aspects of this rhyme are telling. First, the notion that boys and girls, and by extension men and women, embody distinct visual markers to their sex category rings true. Second, the description of social categories that are important, enduring, and meaningful—sex categories—is compelling. Indeed, sex categorization begins early in life, and it relies heavily on visual cues. Moreover, sex categorization exerts considerable influence over subsequent evaluative judgments, leading sociologists to describe sex category as a *master status* (see, e.g., Hughes, 1945). Surprisingly, the basis of children's sex categorizations is not well understood, and most research has focused exclusively on face perception. Here, we examine how preschool children's categorical sex judgments incorporate another cue that reliably distinguishes men from women—body shape.

Men and women differ on a variety of observable attributes. Young children appreciate such cues exhibiting some form of gender knowledge prior to their ability to verbalize it. Before they reach 9 months, infants distinguish faces according to social categories (i.e., sex, Leinbach & Fagot, 1993; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; and race, Kelly et al., 2007). By 12 months, children reliably integrate gendered cues originating in different modalities (i.e., faces and voices; Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). In infancy, sex categories are discernable at a perceptual, if not a conceptual, level.

Young children's ability to label men and women emerges during the 2nd year (Zosuls, et al., 2009). Preschool children rely heavily on superficial cues (i.e., hair and makeup) for sex categorization

---

The authors thank Mike Polito and Casey Levine-Beard for valuable assistance with data collection, Scott Johnson and Diane Ruble for helpful comments on a previous version of this manuscript, Masumi Iida and Jennifer Krull for valuable statistical advice, and the children and their parents for participating in the study. Portions of these data were presented at the 2007 meeting of the Society for Research in Child Development.

Correspondence concerning this article should be addressed to Kerri L. Johnson, Departments of Communication Studies and Psychology, University of California, 2330 Rolfe Hall, Los Angeles, CA 90095. Electronic mail may be sent to Kerri.Johnson@ucla.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0002

(Thompson & Bentler, 1971), even though internal face structure is more reliable (Burton, Bruce, & Dench, 1993). In fact, children are unable to judge the sex of faces devoid of superficial cues until age 7 (Wild et al., 2000). Thus, although young children exploit aspects of appearance that covary with sex to make categorical judgments, they do not utilize sexually dimorphic cues that reliably discriminate men from women. Undoubtedly, children's use of sexual dimorphism for categorization increases throughout development. Yet this process is poorly understood, in part because a preponderance of research focused on face perception.

The importance of the face for person perception is indisputable. In daily life, however, faces are perceived in the context of another sexually dimorphic cue, the body. Furthermore, person perception often occurs from a distance, or visual vantage, that prohibits face processing—forcing perceptions to rely on body cues, many of which are diagnostic of social categories. Both body shape and motion, for example, are sexually dimorphic (Johnson, 2004; Johnson & Tassinary, 2007b; Kerrigan, Todd, & Della Croce, 1998), and adults exploit these differences when making sex categorizations (Lippa, 1983; Pollick, Kay, Heim, & Stringer, 2005). Recent evidence suggests that the body's shape can even be more compelling than its motion for sex judgments (Johnson & Tassinary, 2005).

In children, body knowledge increases with age, first reflecting an understanding of body structure and later revealing a more sophisticated understanding of primary and secondary sex characteristics. Children notice violations of the body's canonical structure quite early (i.e., 15 months; Slaughter & Heron, 2004; Slaughter, Heron, & Sim, 2002; Slaughter, Stone, & Reed, 2004), demonstrating an understanding of the spatial relations among the head, limbs, and torso. By age 3, this knowledge extends to sex-specific structure. Specifically, some children recognize that primary sex characteristics define a target's sex, and can use them for sex categorization even when they are no longer visible and when a target wears cross-sex-typed clothing (Bem, 1989). Under some circumstances, however, superficial cues (e.g., hair) carry more weight in gender labeling. In one study (Thompson & Bentler, 1971), preschool children dressed and labeled dolls as either "Mommy" or "Daddy." These judgments were most heavily influenced by hair length, followed by body anatomy (i.e., secondary sex characteristics), and finally genitals. By age 10, secondary sex characteristics (i.e., body shape) rise in importance and influence more evaluative judgments, such as attractiveness (Connolly, Slaughter, & Mealey, 2004). Thus, young children exhibit increasing knowledge concerning the body's structure, if not an explicit understanding of sexual dimorphism.

In spite of this increase in body knowledge, it is difficult to identify when children notice and use sexually dimorphic cues in social perception. Although some research suggests that young children *can* use primary sex characteristics to inform sex categorization (Bem, 1989), genitals are rarely revealed to others. Instead, observers rely on secondary sex characteristics or adornment cues for sex categorization. To our knowledge, only one study has examined how secondary sex characteristics—the most reliable visual cues under most circumstances—affect sex-typed judgments among children (Thompson & Bentler, 1971), and none has explicitly examined how such cues affect children's sex categorization. Thus, precisely when such cues influence children's judgments remains unclear.

Here we examine whether 4–6-year-old children use body shape, the waist-to-hip ratio (WHR), for sex categorization. This age range reflects a period of rapid changes in children's gender development (Ruble, Martin, & Berenbaum, 2006; Ruble et al., 2007) and increased sensitivity to anatomical sex differences (Thompson & Bentler, 1971). In Study 1, participants completed two tasks—one in which children identified the body that was "most like" a man or woman, and another in which children categorized bodies by sex. In Study 2, we recorded participants' eye movements as they categorized bodies, and we examined visual behavior to discern how children did this.

Study 1

Method

Participants. A total of 73 children (36 boys, 37 girls) between 3.48 and 6.29 years were recruited from an urban community using a commercial database maintained by the department. Four-year-olds ranged from 3.48 to 4.40 years ($M = 4.11$, $SD = 0.21$), 5-year-olds ranged from 4.56 to 5.38 years ($M = 4.97$, $SD = 0.22$), and 6-year-olds ranged from 5.63 to 6.29 years ($M = 6.05$, $SD = 0.17$). Sex distribution was comparable across age group (4-year-olds: 12 boys, 14 girls; 5-year-olds: 10 boys, 9 girls; 6-year-olds: 14 boys, 14 girls). Participants were predominantly Caucasian and came primarily from upper-middle-class backgrounds.

Materials. Stimuli depicted nine computer-generated human bodies that varied in WHR from 0.5 to 0.9 (see Figure 1). Stimuli were rendered in Poser 5™ (Smith Micro Software, Aliso Viejo, CA) using the parameters described in Higa (1999) and Johnson and Tassinary (2007a, 2007b). Bodies were sexually dimorphic in shape but excluded other secondary and tertiary cues to sex (e.g., facial features, body hair, hairstyle, and clothing). These images, and others like them, have been used in research to examine adults' perception of sex categories (Johnson & Tassinary, 2007a, 2007b), attractiveness (Johnson & Tassinary, 2007a, 2007b), and sexual orientation (Johnson, Gill, Reichman, & Tassinary, 2007). These studies found that adult participants perceive the boundary between "female" and "male" bodies to be a WHR of 0.7 (Johnson, 2004; Johnson & Tassinary, 2009). Each image was mounted (separately) on card stock and laminated for durability.

Procedures. Participants were tested individually. We familiarized participants with all stimuli by showing all cards simultaneously, increasing in WHR. We asked participants to identify how the images differed. After the participant indicated that the images differed in body shape (most frequently by pointing to the waist and hips region) participants completed two tasks, in counterbalanced order.

In an identification task, all targets were presented simultaneously. Cards were mounted on a wall, increasing in WHR. Participants stood 2 ft away from the stimuli, and the experimenter drew the participant's attention to the full range by gesturing. Then participants identified which body was "most like a man" and "most like a woman" by placing a sticker on the chosen body. We recorded the corresponding WHR for each selection, which served as the dependent variable in our analysis of these data.

In a categorization task, participants categorized each target's sex by sorting cards into boxes depicting the face of either a man or a woman (left or right placement of boxes was counterbalanced across participants). Participants practiced the procedure by sorting cartoon images (four *Peanuts* characters: Charlie Brown, Linus, Lucy, and Sally) by sex. Then, participants categorized each of the nine targets twice, in two randomized sets.

Results and Discussion

We analyzed data using a generalized equation estimation technique (GEE; Diggle, Liang, & Zeger, 1994; Fitzmaurice, Laird, & Ware, 2004). This permitted us to estimate within-subject parameters for both binary and continuous dependent variables for a fully within-subject factorial design. We use a standard regression vernacular, however, to describe our results. We centered age in years around its mean (i.e., 5.06), and participant sex, and WHR (in integers) at zero. We coded judged sex numerically (0 = *male*, 1 = *female*). We report parameter estimates as unstandardized regression coefficients for all significant effects in Table 1. Task order was initially included as a factor in analyses. It did not affect any dependent variable, and will receive no further mention.

Identification task. We examined whether the bodies identified to be "most like a man" and "most like a woman" varied as a function of age and sex. Thus, the selected WHR was our dependent variable. We regressed WHR onto judgment (i.e., most like man or most like a woman), participant age, participant sex, and all interactions. As predicted, bodies judged to be most like a man had larger WHRs (i.e., .11 units) than bodies identified to be most like a woman, and this varied by age (see Table 1).

We followed up the judgment by age interaction by testing the simple slopes of judgments centered (separately) at each age (see Figure 2 and Table 1). Among 4-year-olds, judgments did not vary significantly with WHR. Among 5- and 6-year-olds, the

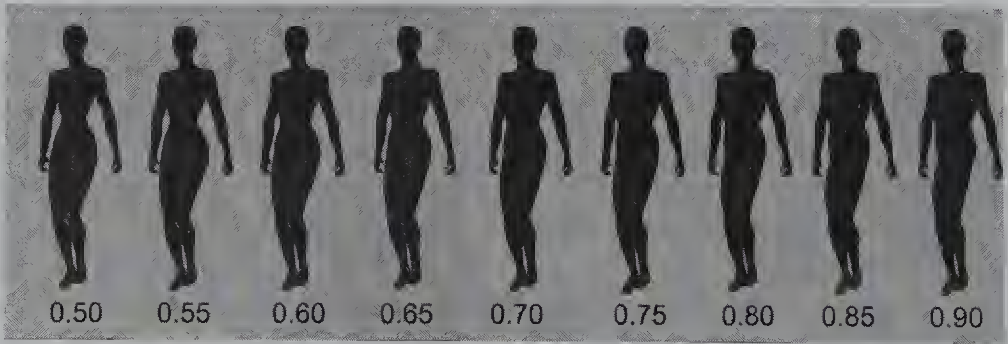


Figure 1. Stimuli depicted nine computer-generated bodies that varied in waist-to-hip ratio from 0.5 to 0.9.

Table 1

Summary of Generalized Equation Estimates (GEE) Predicting Judgments in the Identification and Categorization Tasks of Study 1

Effect	*B*	*SE*	*z*	*p*	Odds ratio
Identification task					
Man/woman	−0.1137	0.021	−5.48	****	N.A.
Interaction (Age × Man/Woman)	−0.0768	0.0219	−3.51	***	N.A.
Simple slope at age 4	−0.0321	0.0289	−1.11	*ns*	N.A.
Simple slope at age 5	−0.1088	0.0206	−5.29	****	N.A.
Simple slope at age 6	−0.1856	0.0311	−5.96	****	N.A.
Categorization task					
WHR	−0.0947	0.0218	−4.33	****	2.35
Interaction (Age × WHR)	−0.0908	0.0252	−3.61	***	—
Simple slope at age 4	0.0019	0.0320	0.06	*ns*	1.02
Simple slope at age 5	−0.0890	0.0217	−4.10	****	2.23
Simple slope at age 6	−0.1798	0.0344	−5.23	****	5.04
Sex of participant	0.5077	0.2045	2.48	**	1.66

Note. B values are unstandardized coefficients from GEE. Odds ratio was computed based on the change in judgments of the full range of waist-to-hip ratio (WHR) stimuli (i.e., 0.5–0.9). Nonsignificant main effects and interactions are omitted.
$^{**}p \leq .01$. $^{***}p < .001$. $^{****}p < .0001$.

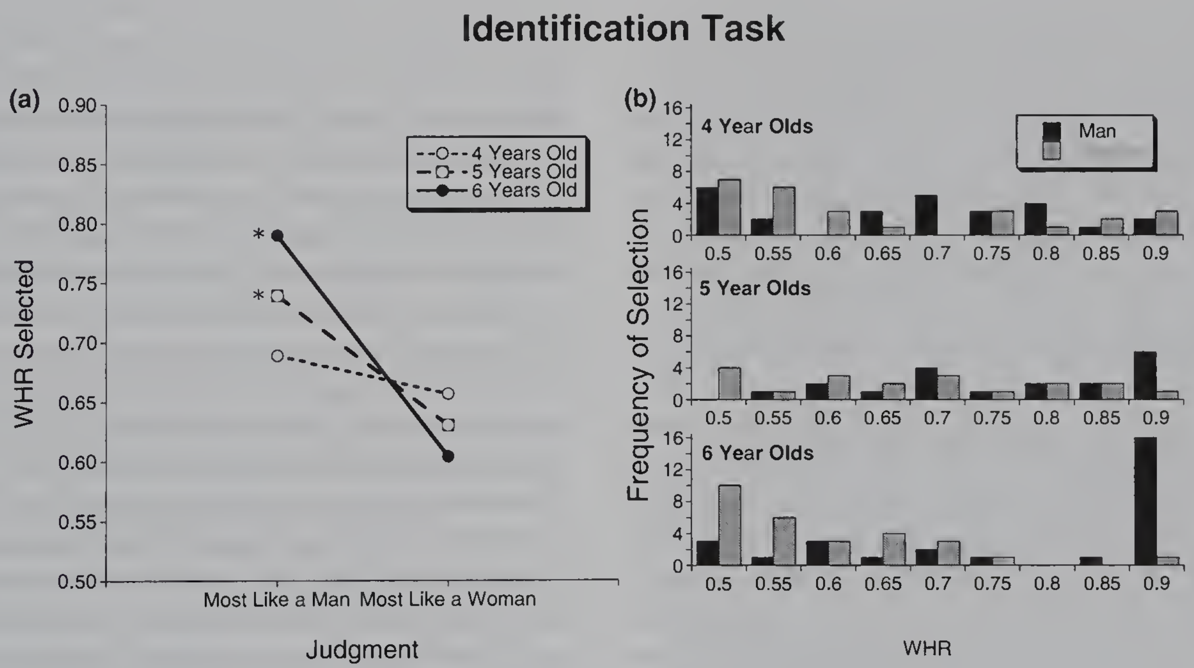


Figure 2. Waist-to-hip ratio (WHR) selected to be "most like a man" and "most like a woman" as a function of age for the identification task. (a) Plots reflect predicted WHR selected for each judgment centered separately around 4, 5, and 6 years of age. (b) Raw frequency that each WHR was selected.

WHR identified to be most like a man was significantly larger than the WHR identified to be most like a woman, simple. Other effects were nonsignificant, all *z*s < 0.91, *ns*.

Categorization task. First, we examined the accuracy of sex categorizations during the practice phase. No errors occurred at any age. Because participants sorted cards twice, we initially included this as a factor in all analyses. No main effects or interactions involving set approached significance (all *B*s < 0.09, all *p*s > .34), and it was therefore dropped from the analyses.

To examine whether sex categorizations varied with age and WHR, we regressed perceived sex onto WHR, participant age, participant sex, and all interactions. As predicted, as the WHR increased, targets were more likely to be categorized as men, and this varied by age (see Table 1).

We followed up the WHR × Age interaction by testing the simple slopes of judgment centered

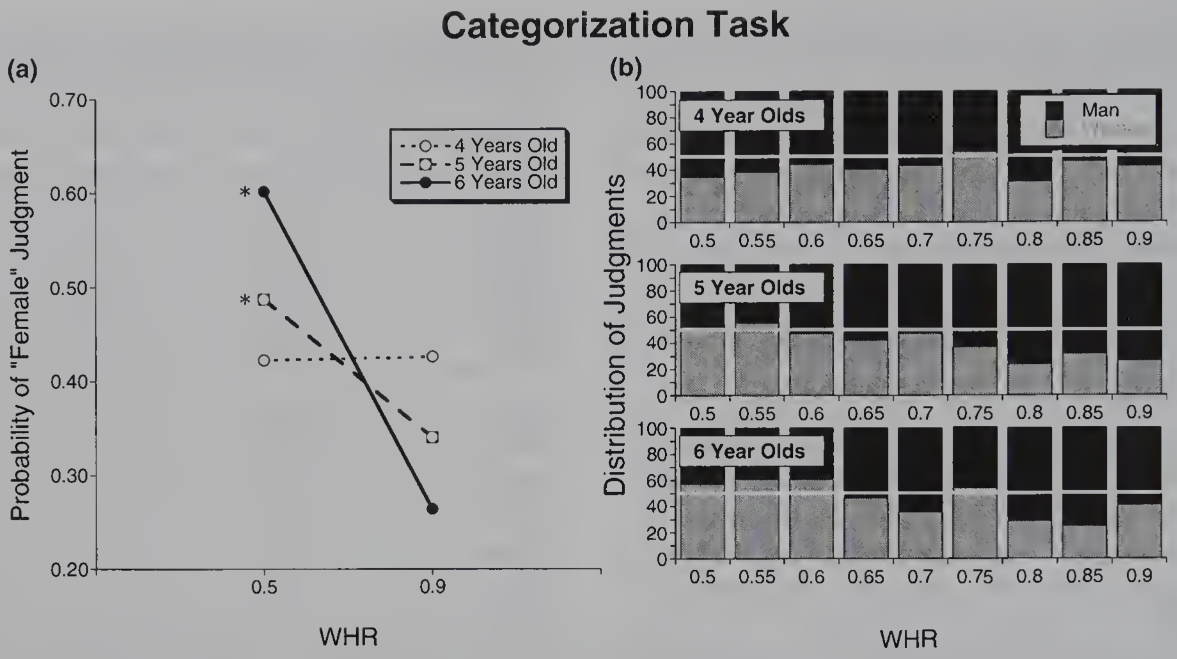


Figure 3. Probability that a target is judged to be female as a function of waist-to-hip ratio (WHR) and age for the categorization task. (a) WHR was treated as a continuous variable in this analysis, and plots reflect predicted probability when the WHR was 0.5 and 0.9, centered separately around 4, 5, and 6 years of age. (b) Raw frequency of male and female categorizations for each WHR.

(separately) at each age (see Figure 3 and Table 1). Among 4-year-olds, judgments did not vary with WHR. Among 5-year-olds, each increase in WHR corresponded to 8.9% more "man" categorizations; among 6-year-olds, each increase in WHR corresponded to 17.98% more "man" categorizations.

Compared with girls, boys were more likely to categorize targets to be men and vice versa (see Table 1). No other effects reached significance, all zs < 1.25, *ns*.

These results suggest that between 4 and 6 years, children increasingly exploit sexual dimorphism for sex categorization. This pattern could occur for many reasons. One possibility is that our tasks were too difficult for 4-year-olds. This seems unlikely. First, 4-year-olds easily and accurately categorized each practice stimulus. Although the demands of each task differed (e.g., relative judgments in the identification task, sorting in the categorization task), the results from both were convergent, suggesting that task difficulty is unlikely to be the explanation. Instead, we propose that 4-year-olds' failure to utilize WHR reflects a lack of knowledge about sexual dimorphism. This could contribute to their failure to utilize the WHR in their judgments in two ways. First, they may not attend to the waist and hips because they do not perceive the region as relevant for sex categorization. Second, they may attend to differences in WHR but fail to relate those differences systematically to sex categorization. These issues concerning younger children's ability to identify and utilize a sexually dimorphic cue cannot be resolved in Study 1. In Study 2, we address these issues by observing children's visual behavior as they categorized our stimuli.

The utilization of the WHR for sex category judgments corresponds to telltale visual behaviors. Using similar stimuli, Johnson and Tassinary (2005) recorded adult participants' eye movements as they visually scanned bodies prior to making social judgments. When participants were required to make sex category judgments, they concentrated visual scanning in the waist and hips, recognizing the region's relevance for the task. In contrast, when the targets' sex was prespecified, thus removing the need to categorize, participants no longer concentrated scanning within the waist and hips. Such measures are also likely to be informative for understanding how children come to use sexual dimorphism for sex categorization. We reasoned that observing children's visual behavior would allow us to determine not only whether children recognize differences across the stimuli but also whether they recognize the region's relevance for their judgments.

Study 2

In Study 2, we tracked children's eye movements as they made sex category judgments. First, we calculated visual scanning within the waist and hips as an index of children's attention toward the relevant cue and examined whether this differed by age. Then we examined whether attention toward this area predicts its use for sex categorization. We predicted that children who look toward the waist and hips would be more likely to show a strong relation between WHR and sex category judgments.

Method

Participants. Participants were forty-one 4- and 6-year-old children (20 boys and 21 girls). The 4-year-olds ranged from 3.73 to 4.25 years (*M* = 3.95, *SD* = 0.12); the 6-year-olds ranged from 5.67 to 6.52 years (*M* = 6.05, *SD* = 0.17). Participants were recruited as described previously. Sex distribution was comparable across age groups (4-year-olds: 10 boys and 10 girls; 6-year-olds: 10 boys and 11 girls).

Procedures. As before, we familiarized participants with the range of stimuli by showing all targets simultaneously, increasing in WHR. After indicating that they perceived differences between the targets, participants completed a computerized categorization task.

Participants were seated approximately 120 cm away from a 43 cm Tobii 1750 (Falls Church, VA) corneal reflection eye-tracker. ClearView software (Tobii Software, Falls Church, VA) presented stimuli and recorded eye movements at 30 Hz, with a spatial resolution within .5° visual angle. Participants were calibrated individually by registering five known fixation points, and confirming their reliability before proceeding.

Stimuli included digitized versions of those used in Study 1 (see Figure 4), and the task was conceptually similar, albeit with virtual rather than real boxes. In this task, the "boxes" appeared as photographs of a man and woman, counterbalanced in order, in the upper corners of the screen. The task was described using an identical script. That is, we instructed participants to determine which box each stimulus belonged in. We familiarized participants with the task—first verbally, then with a practice phase as in Study 1. Then, participants categorized the nine targets twice, in one of two random orders.

Results and Discussion

Preliminary analysis. No participant reported any confusion about the task, and no participant made an error in categorizing the practice stimuli. We confirmed that the basic pattern of judgments replicated using the same analyses described in Study 1. Sex category judgments varied, albeit marginally, as a function of age and WHR, interaction $B = -0.0830$, $SE = 0.0475$, $z = -1.75$, $p = .08$. Among 4-year-olds, judgments did not vary with WHR, simple $B = 0.2600$, $SE = 0.2284$, $z = 1.14$, *ns*. Among 6-year-olds, sex category judgments did vary with WHR, $B = -0.5949$, $SE = 0.2689$, $z = -2.21$, $p = .0269$, odds ratio (OR) = 5.04 (predicted means for WHR of 0.5 = 65% female judgments, and for WHR of .9 = 21% female judgments).

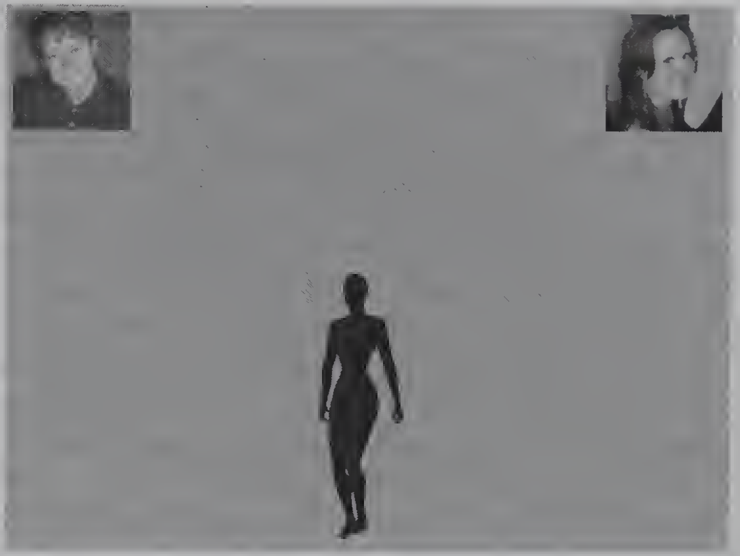


Figure 4. Sample stimulus from Study 2.
Note. The stimulus was centered in the lower portion of the screen; the "male box" and the "female box" appeared in the upper corners (their location was counterbalanced across participants).

Attention to WHR. First, we computed the proportion of body scanning directed toward the waist and hips (hereafter, *Dwell*). Adults tend to increase scanning within the waist/hip when they categorize bodies by sex (Johnson & Tassinary, 2005), but not when the sex of a target is already known. We predicted that a similar difference in scanning would emerge when comparing the scan patterns of children who recognize the relevance of body shape for sex categorization to those who do not. Thus, we suspected that scanning this region might differ with age. Indeed, older children, relative to younger children, looked significantly more to the waist and hips of each body, $B = 0.0368$, $SE = 0.0191$, $z = 1.93$, $p = .05$ (*M*s = 15% and 23%). Importantly, we found no age differences in scanning of the head, chest, or legs (all *B*s < 0.033, all *z*s < 1.5, *ns*). This pattern, combined with the age differences observed in the effect of WHR for categorization, is consistent with our hypothesis that younger children do not fully recognize the relevance of this region as a cue to sex category membership.

Next, we tested this more directly by examining whether visual scanning of the waist and hips predicted children's use of the WHR for their judgments. We used hierarchical linear modeling for this analysis (HLM; Raudenbush, Bryk, & Congdon, 2007). Parameter estimates are shown in Table 2. At Level 1, the analysis modeled the average relation between WHR and judgment. Overall, the two were significantly related. This replicated the overall effect of WHR on judgments that we found in Studies 1 and 2. At Level 2, this analysis examined whether Dwell predicted the strength of the Level 1 relation between WHR and judgments. Not surprisingly, Dwell did not predict judgment directly, but it did moderate the relation between WHR and judgment. As seen in Figure 5, when scanning within the waist and hips was low, the relation between WHR and judgment was negligible. When scanning within the waist and hips was high, in contrast, the relation between WHR and judgment was strong. Importantly, this effect did not differ by participant age when it was included as a Level 2 variable, and scanning within other body regions did not predict the relation between WHR and judgments (all *t*s < 1.81, *ns*).

These results indicate that, as in adults (Johnson & Tassinary, 2005), children's visual scanning of a sexually dimorphic region of the body is functional. Doing so informs sex categorization. Children with elevated visual scanning within the waist and hips showed a strong relation between WHR and judgment; children with no elevated scanning in this area did not. Importantly, this basic effect was not moderated by age. This suggests that younger children's lesser scanning of a sexually dimorphic region of the body occurs because they are less likely to recognize its relevance as a cue for the task at hand—sex categorization—not because they fail to appreciate the differences in the first place. This is important in two ways. First, these findings suggest that the age cohort effects in Study 1 likely reflected a difference in utilization, rather than a wholesale failure of younger children to notice differences between stimuli. Second, this suggests that the use of sexual dimorphism for categorization occurs between ages 4 and 6.

Table 2

Summary of Hierarchical Linear Modeling (HLM) Estimates Predicting Judgments in Study 2

Effect	*B*	*SE*	*t* (*df*)	*p*
Level 1				
WHR	−3.72	1.21	−3.08 (39)	**
Level 2				
Dwell within waist and hips	−1.12	0.79	−1.41 (39)	*ns*
Interaction (Dwell × WHR)	−20.31	8.99	−2.26 (39)	*

Note. *B* values are unstandardized coefficients from hierarchical linear modeling (Raudenbush et al., 2007). WHR = waist-to-hip ratio.
$^{*}p < .05$. $^{**}p < .01$.

General Discussion

These findings begin to fill a gap in the literature that has received minimal attention—how children utilize sexually dimorphic cues in social categorization. In Study 1, we found that between the ages of 4 and 6, children's sex category judgments increasingly covaried with a sexually dimorphic cue, the WHR. Among 5- and 6-year-olds, but not 4-year-olds, targets were more likely to be categorized *male* as the WHR increased. Additionally, across this age range, the WHRs of bodies identified to be most like a man and most like a woman became increasingly divergent, reaching significance by age 5. In Study 2, we demonstrated that increased scanning of this region predicted stronger associations between WHR and judgments. Collectively, these findings highlight a developmental shift between the ages of 4 and 6 in which children become adept at exploiting morphological differences to disambiguate sex category membership.

Because our stimuli were computer generated, we should remain mindful of two things. First, in day-to-day social perception, children's judgments are likely to be informed by multiple visual cues, some of which may take priority over perception of the body. Second, our findings may reveal more about children's judgments of representational media than the actual perception of others. In spite of this, other research suggests that media portrayals of men and women are routinely incorporated into judgments, at times biasing perceptions and preferences (see, e.g., Aubrey & Harrison, 2004; Yu & Shepard, 1998). Therefore, any effects that emerged as a result of the representational nature of our stimuli are likely to be incorporated into perceptions of others as well.

Assuming that our results indicate something general about how children perceive others, the age at which children begin to use sexual dimorphism for sex categorization occurs at a point of rapid transitions in gender development. The increasing awareness that certain visual cues are reliable indicators to a target's sex may demarcate a transition of functional significance—a period during which children shed the notion that one's sex is mutable. Indeed, between 4 and 6 years, young children come to understand the permanence of their own

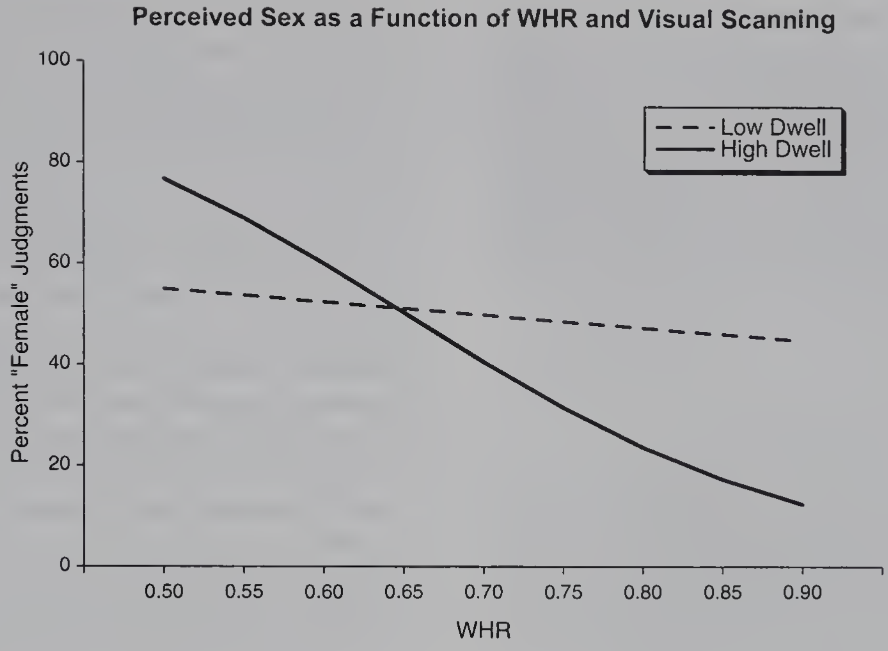


Figure 5. Probability that a target is judged to be female as a function of waist-to-hip ratio (WHR) and visual scanning behavior. *Note.* Plots reflect predicted probability at each WHR, centered separately around low and high Dwell.

sex (gender constancy), a phenomenon researchers have linked to increases in gender stereotype knowledge (Martin, Ruble, & Szkrybalo, 2002; Ruble et al., 2006). An increased reliance on cues that are biologically diagnostic of a target's sex, as opposed to those that are merely correlated with sex, for sex categorization may undergird gender constancy.

Importantly, others have described a functional relation between body knowledge and gender constancy, albeit at a much younger age. Bem (1989) notably found that children who used genitals as the criterion for sex categorization also exhibited some forms of gender constancy by age 3. These findings underscore a minimal age at which children *can* exhibit rudimentary forms of gender constancy, but such effects are likely to be constrained in the real world. In social interactions, for example, genitals are rarely displayed to compel unquestionable categorizations. Instead preschool children are prone to use superficial cues to appearance for categorization (Thompson & Bentler, 1971), and their judgments tend to fluctuate with shifting evidence (e.g., when a boy dons a dress). Malleable sex categorization may be particularly likely when (a) studies use stimuli depicting prepubescent children who are devoid of secondary sex characteristics, and/or (b) children are not equipped to incorporate the dimorphic cues that are visible in stimuli depicting adults. In both cases, children's judgments are more probabilistic than definitively categorical. As children develop an appreciation that sexually dimorphic body cues are visible for fully clothed targets, however, their judgments may shift from reflecting malleable and probabilistic social perceptions to reveal stable categorizations that are less influenced by superficial changes in appearance.

The coincidence of age at which children utilize sexually dimorphic body shape for categorization and at which gender constancy emerges dovetails nicely with research in essentialism. Essentialism refers to the perception that visible cues reveal underlying essential differences between categories (see, e.g., Gelman, 2003). Much of the research in gender essentialism has focused on the perception that stereotyped behaviors emerge from biological rather than social origins, yet the development of such beliefs is informative herein as well. Specifically, we examined young children's ability to infer categories from visual cues (see, e.g., Gelman, Collman, & Maccoby, 1986). From this perspective, children's emerging ability to use sexual dimorphism for sex categorization may reveal a growing awareness of biological differences between men and women and an appreciation that some cues reliably differentiate the two. This reasoning predicts children's increased reliance on sexually dimorphic cues and an emphasis on those cues over superficial aspects of appearance for sex category judgments. Indeed, we found an increased reliance on sexual dimorphism for sex categorization by age 5. The reliance on sexually dimorphic body shape precedes the age at which children's categorization of faces favors sexual dimorphism over superficial cues such as hairstyle (i.e., 5 vs. 7 years). This difference could be because we were seeking the age when children recognize the relevance of sexual dimorphism, whereas prior research in the categorization of faces sought to establish at what age children favor sexually dimorphic cues over superficial appearance cues (Wild et al., 2000). In addition to the research in face perception and gender constancy, our findings provide convergent evidence for the development of essential concepts of men and women in early childhood.

Conclusion

The question of how observers categorize others according to social categories has been an enduring interest of both artists and scientists alike. Although our article opened with an artistic portrayal of the essential characteristics that distinguish boys from girls, our empirical work was decidedly more focused on the perception of a cue that is inherently sexually dimorphic. As such, our findings represent a next step in understanding the development of social categorization—by specifying the age at which children's sex category judgments reflect sensitivity to sexually dimorphic body shape.

References

Aubrey, J. S., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, *6*, 111–146.

Bem, S. L. (1989). Genital knowledge and gender constancy in preschool children. *Child Development*, *60*, 649–662.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153–176.

Connolly, J. M., Slaughter, V., & Mealey, L. (2004). The development of preferences for specific body shapes. *Journal of Sex Research*, *41*, 5–15.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford, UK: Oxford University Press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New York: Wiley.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *57*, 396–404.

Higa, M. (1999). *Perception based character modeling and animation*. Unpublished master's thesis, Texas A&M University, College Station.

Hughes, E. C. (1945). Dilemmas and contradictions of status. *American Journal of Sociology*, *50*, 353–359.

Johnson, K. L. (2004). Perceiving people: Interpersonal meaning in the body's motion and morphology. *Dissertation Abstracts International*, *65*, 2146.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology*, *93*, 321–334.

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science*, *16*, 890–897.

Johnson, K. L., & Tassinary, L. G. (2007a). Compatibility of basic social perceptions determines perceived attractiveness. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 5246–5251.

Johnson, K., & Tassinary, L. G. (2007b). Interpersonal metaperception: The importance of compatibility in the aesthetic appreciation of bodily cues. In V. Swami & A. Furnham (Eds.), *Body beautiful: Evolutionary and sociocultural perspectives* (pp. 159–184). New York: Palgrave Macmillan.

Johnson, K. L., & Tassinary, L. G. (2009). *Person (mis)perception: Presumed and erroneous extremity in the categorization of bodies*. Unpublished manuscript, University of California, Los Angeles.

Kelly, D. J., Shaoying, L., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., Liu, Q., & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy*, *11*, 87–95.

Kerrigan, D. L., Todd, M. K., & Della Croce, U. (1998). Gender differences in joint biomechanics during walking: Normative study in young adults. *American Journal of Physical Medicine and Rehabilitation*, *77*, 1–7.

Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development*, *16*, 317–332.

Lippa, R. (1983). Sex typing and the perception of body outlines. *Journal of Personality*, *51*, 667–682.

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*, 903–933.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1247–1265.

Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, *30*, 436–442.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*, 1109–1121.

Raudenbush, S., Bryk, A., & Congdon, R. (2007). HLM6: Hierarchical Linear and Nonlinear Modeling (Version

6.04) [Computer software]. Lincolnwood, IL: Scientific Software International.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 3, 6th ed., pp. 858–932). New York: Wiley.

Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, *78*, 1121–1136.

Slaughter, V., & Heron, M. (2004). Origins and early development of human body knowledge. *Monographs of the Society for Research in Child Development*, *69*(2).

Slaughter, V., Heron, M., & Sim, S. (2002). Development of preferences for the human body shape in infancy. *Cognition*, *85*, B71–B81.

Slaughter, V., Stone, V. E., & Reed, C. (2004). Perception of faces and bodies: Similar or different?. *Current Directions in Psychological Science*, *13*, 219–223.

Thompson, S. K., & Bentler, P. M. (1971). The priority of cues in sex discrimination by children and adults. *Developmental Psychology*, *5*, 181–185.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*, 55–75.

Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, *77*, 269–291.

Yu, D. W., & Shepard, G. H., Jr. (1998). Is beauty in the eye of the beholder? *Nature*, *396*, 321–322.

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed plays. *Developmental Psychology*, *45*, 688–701.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1356–1366

Cognitive Flexibility in Drawings of Bilingual Children

Esther Adi-Japha, Jennie Berberich-Artzi, and Afaf Libnawi
Bar-Ilan University

A. Karmiloff-Smith's (1990) task of drawing a nonexistent object is considered to be a measure of cognitive flexibility. The notion of earlier emergence of cognitive flexibility in bilingual children motivated the current researchers to request 4- and 5-year-old English–Hebrew and Arabic–Hebrew bilingual children and their monolingual peers to draw a flower and a house that do not exist ($N = 80$). Bilinguals exhibited a significantly higher rate of interrepresentational flexibility in their drawings (e.g., "a giraffe flower," "a chair-house," found in 28 of 54 drawings), whereas the level of complex intrarepresentational change was similar across groups. Interrepresentational drawings were previously reported only for children older than 7 years. The specific mechanisms by which bilinguals' language experience may lead to interrepresentational flexibility are discussed.

As children develop, they learn to adjust to changing demands and priorities, to consider something from a fresh or different perspective, to switch between perspectives, and to "think outside the box." These behaviors are referred to as "cognitive flexibility" and build on inhibitory control mechanisms and working memory, which are together regarded as core executive skills critical for cognitive development (Diamond, Barnett, Thomas, & Munro, 2007). Current research shows that early bilingualism leads to a developmental advantage in nonlinguistic tasks that necessitate cognitive flexibility (e.g., Bialystok & Senman, 2004; Carlson & Meltzoff, 2008). This advantage exists as early as age 4 (e.g., Bialystok, 1999; Bialystok & Martin, 2004) and is maintained in adulthood (Bialystok, Craik, & Ryan, 2006; Costa, Hernandez, & Sebastian-Galles, 2008).

For over half a century, drawing has been used to study cognitive development (Arnheim, 1956; Freeman, 1995; Golomb, 1974, 1992; Luquet, 1927; Piaget & Inhelder, 1948/1967). Among the variety of drawing tasks that tap into executive skills, Karmiloff-Smith's (1990) task of drawing nonexistent objects is considered to be a measure of cognitive flexibility (Spensley & Taylor, 1999), or representational flexibility (Picard & Vinter, 1999, 2006, 2007), present at that age. Children's skilled graphic production implicates procedural as well as representational aspects. The task allows for few distinct modification categories that involve the relaxation of procedural and representational constraints (Karmiloff-Smith, 1990; Picard & Vinter, 2007). The current study compared the performance of monolingual and bilingual children whose control over specific inhibitory control mechanisms was shown to differ (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008).

In the drawing of nonexistent objects task children are asked to draw "an X that does not exist," such as an X they invent, that they have never seen before, a strange X, an X with something funny or odd about it (Karmiloff-Smith, 1990; Spensley & Taylor, 1999). Karmiloff-Smith (1990) showed that in response to a request to draw a man "who does not exist" (and the same for an animal and a house), 4- to 6-year-olds tended to make size, shape, or deletion changes to elements within their drawings, whereas 8- to 10-year-olds exchanged the position of elements, added extra elements from the same representation (e.g., a man with two heads), or included cross-category insertions that involved synthesizing two different representations (e.g., a house with wings).

Different studies that replicated Karmiloff-Smith's (1990) original findings (Berti & Freeman, 1997; Spensley & Taylor, 1999; Zhi, Thomas, & Robinson, 1997) verified a classification of these modifications into two main types of representational flexibilities: An interrepresentational flexibility

---

The authors thank Professor Norman Freeman and Dr. Gil Deisendruck for comments on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Esther Adi-Japha, School of Education, Bar-Ilan University, Ramat-Gan 52900, Israel. Electronic mail may be sent to japhae@mail.biu.ac.il.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0003

observed when a link between components from different categories appears in one drawing (as displayed by cross-category insertions), which emerges at around 8 years, and an intrarepresentational flexibility connected to changes seen in the components of a given graphic representation (as displayed by the other types of modifications), seen already in 4- and 5-year-olds.

Although 4- and 5-year-olds typically did not spontaneously produce interrepresentational changes or complex intrarepresentational changes (i.e., position or orientation change and same-category insertion), they did so when explicitly requested: for example, when asked to draw an animal-man (i.e., a man who is both an animal and a man) and a house-man (Berti & Freeman, 1997); when asked to draw a house with wings or a mushroom with arms (Picard & Vinter, 2007); when given an earlier specific example, for example, a man with wings instead of arms when the task was to ''draw a man with some parts being replaced with parts of an animal'' (Spensley & Taylor, 1999); or when shown a picture they could follow (i.e., a woman with two heads; Zhi et al., 1997). Furthermore, in Spensley and Taylor's (1999) experiment, 4- to 6-year-old children succeeded in drawing examples that were specified verbally, but they were unable to utilize this ability independently. These studies suggest a developmental constraint: Children who are less than 7 years of age cannot respond to the task ''draw an X that does not exist'' with complex representational change without being primed by an explicit example that they can follow carefully, or without being instructed to produce a specific type of insertion.

Further experiments verified that complex intrarepresentational as well as interrepresentational modifications were associated with procedural flexibility (Hollis & Low, 2005; Picard & Vinter, 1999, 2007). However, the notion that interrepresentational flexibility crucially requires the relaxation of a constraint of independence between representations may suggest different pathways to cognitive flexibility for complex intrarepresentational and interrepresentational change in terms of the inhibitory control processes implied.

In this study, drawing of nonexistent objects made by 4-year-old English–Hebrew and 5-year-old Arabic–Hebrew bilingual children were compared with those made by their Hebrew-speaking monolingual peers. Representational innovations made by 4-year-olds are of special interest because at this age children just master their representational skills (Thomas & Silk, 1989). Previous studies of 4-year-old bilingual children verified their advantage in tasks that tap cognitive flexibility (Bialystok, 1999; Bialystok & Martin, 2004). In our pilot studies, we saw that a man, a house, or an animal were too difficult for children aged around 4, whereas a flower was reasonably drawn by most children of that age. At the age of 5 years we saw that although a house was reasonably drawn, an animal was still too difficult to draw, and when these children were requested to draw a nonexistent man, some irrelevant responses (e.g., a dead man) emerged. In Study 1, 4-year-old English–Hebrew bilinguals and their monolingual Hebrew-speaking peers were therefore asked to draw a flower and a flower that does not exist. In Study 2, 5-year-old Arabic–Hebrew bilinguals and their monolingual Hebrew-speaking peers were also asked to draw a house and a house that does not exist. The Arabic–Hebrew bilinguals were tested as a second bilingual group and to control for the possibility that the different direction of writing (right-to-left in both Arabic and Hebrew, as opposed to English) may contribute to the grapho-motor flexibility in the English–Hebrew bilinguals' production. If indeed bilinguals are more successful in drawing nonexistent objects, this will comprise further support for the notion that flexibility in the linguistic domain may facilitate flexibility in other nonlinguistic domains.

Study 1

Method

Participants

Twenty-five monolingual children (12 boys and 13 girls) and 25 bilingual children (13 boys and 12 girls) participated in this study. The participants were 4-year-olds (range = 42–60 months, $M = 49.76$, $SD = 3.79$ months and $M = 51.12$, $SD = 4.75$ months for the monolingual and bilingual children, respectively). The monolingual children attended a private Hebrew kindergarten. The bilingual children attended a private English–Hebrew bilingual kindergarten. The kindergartens were in the same neighborhood, which is located in an upper-middle-class Israeli suburb, and operated for at least 9 hr a day.

For the bilingual group, kindergarten teacher reports ensured that the children participating in the study had acquired their English from home. In this group, either both parents were native English speakers who spoke both English and Hebrew at home and in the community, or one parent spoke English and the other spoke Hebrew.

Procedure

Children were observed individually in a quiet room at their kindergarten. Testing took place in 1–2 sessions (session time did not exceed 20 min). Bilingual children had an additional session for the English vocabulary test. Summer drawing always preceded the nonexistent flower task (see next). The experiment took place in 2 successive years with similar results.

Measures

Peabody Picture Vocabulary Test–III. The Peabody Picture Vocabulary Test–III (PPVT–III; Dunn & Dunn, 1997) is a test of receptive vocabulary. Bilingual children were tested using the English PPVT–IIIA. Following the general procedures of Bialystok (1988), the monolingual and bilingual children were tested on an expert translation of the PPVT–IIIB developed at Bar-Ilan University (see also Bialystok, Shenfield, & Codd, 2000). Different forms of the test were used such that there was no overlap in the specific vocabulary tested. Testing continued until the child made 8 errors out of a block of 12 trials. Raw scores are reported.

The visual reception test. This test was taken from the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, Mccarthy, & Kirk, 1968) and measures the ability to infer meaning from pictures. During the test, the child is presented with a picture for 3 s. On the following page, the child is requested to identify one of four pictures, which resembles the image on the previous page.

Drawing summer. The children were asked to draw the summer. This drawing was used to test their free drawing skills. Drawing complexity was measured using the Kellogg Scale (see next).

''Draw a flower that does not exist'' test. The children were first asked to draw a flower. After completing their drawing, they were requested to draw ''a flower that does not exist.'' Several phrasings were used to enable the children to understand the task: A flower you invent, you have never seen before, a strange flower, with something funny/odd, something make-believe, pretend. After drawing the nonexistent flower, the children were asked to verbalize why such a flower does not exist. Following the procedure developed by Karmiloff-Smith (1990), two independent judges scored the categories of changes as no change, deletion of elements, insertion of new (same-category) elements, change in element shape or size, whole-shape changes, position or orientation changes, and cross-category insertions. Cohen's κ coefficients for interrater agreement were above .9 ($p < .001$) in all change categories. Disagreements were settled by discussion. As in the original study, the categories were not mutually exclusive.

The drawing was classified as showing change if the nonexistent flower violated ''flowerness.'' Similar flower and nonexistent flower, a bigger and smaller flower, or a flower and a different flower were classified as not showing change, even if the verbal response indicated some difference (e.g., this flower does not exist because ''something shines on it'' or ''because it is green, in make-belief'').

The drawing complexity scale. The Kellogg Scale was used to score the complexity of the drawings of the summer, flower, and nonexistent flower. This scale has seven figural prerepresentational complexity levels (Kellogg, 1970; Golomb, 1992): (1) *scribbling*; (2) placing scribbling marks in *patterns*; (3) scribble constellation that constitutes *emergent diagrams* in the shapes of circles, ovals, triangles, rectangles, or crosses; (4) six outlined figures comprise the stage of *diagrams*—the rectangle, the oval, the triangle, the Greek cross, the diagonal cross, and the odd shape that is closed but irregular; (5) shapes that incorporate two diagrams are termed *combines*; (6) shapes that incorporate three or more diagrams are termed *aggregates*; and (7) complex graphic formulas joined into figures that are devoid of pictorial intention or likeness to an object. Three more levels were added because the tasks used in this study explicitly requested drawings that resemble a known object (representations corresponding to summer or flower) that incorporate a relatively simple scheme (at least at the complexity of a combine, as in item 5). These levels include: (8) a recognizable figure composed of two line objects (e.g., in the flower drawing a center and petals; a stalk and an outline of a flower or circle); (9) a recognizable figure composed of three line objects (e.g., in the flower drawing a stalk, a center, and petals); and (10) a recognizable figure that includes complex graphic formulas. For the summer drawing, the most complex element in the drawing was scored. The two independent judges who scored the drawings for the change categories also scored the drawings for complexity. The scores ranged between 2 and 10. Interjudge agreement was, on average, .96. Cohen's κ coefficients for interrater agreement on a categorical scale were, on average .73 ($p < .001$), and weighted

κ coefficients for an ordinal scale were, on average .92 ($p < .001$). Kellogg Scale scores of the three drawing tasks were significantly correlated, $r(50) > .75$, $p < .001$, and significantly correlated with age $r(50) > .51$, $p < .001$.

Results and Discussion

Monolingual and bilingual children performed similarly on the PPVT–III test, and had similar visual reception and drawing complexity scores (Table 1). Furthermore, the bilingual group had comparable language skills in Hebrew and in English, $t(24) = 1.73$, *ns*. Although similar language skills in monolingual and bilingual young children are not common, similar findings were reported in other studies (e.g., Davidson, Jergovic, Imami, & Theodos, 1997).

As in the original study, we used six change categories (see Method), with a slight difference: Changes in the position of flower elements did not appear. Change categories were not mutually exclusive (Karmiloff-Smith, 1990). Drawings were classified as having 0–4 category changes with most children (45/50) producing 0–2 changes (Picard & Vinter, 2007).

Table 2 presents the number of children in each of the "object that does not exist" categories of change (Karmiloff-Smith, 1990). Monolinguals' most frequent change was deletion (Figure 1), whereas bilinguals' most frequent change was cross-category insertions (Figure 2). Differences between groups in these categories were significant —for deletion, $\chi^2(1) = 6.35$, $p < .05$, and for cross-category insertions, $\chi^2(1) = 11.53$, $p < .001$—although deletion results should be interpreted with some caution given that multiple chi-square tests were performed without making an adjustment in the alpha level. The 13 cross-category insertions made by the bilingual children consisted of modified flowers that included human features (7 children) and those that included animal features (6 children).

As can be seen from the figures, children's verbal response helped explain the drawings' content (Adi-Japha, Levin, & Solomon, 1998). The verbal description matched production in 43 of the 50 children who participated in the study (22 monolinguals and 21 bilinguals). One child did not respond. Overall, the drawing descriptions of 13 bilingual versus 2 monolingual children were of a cross-category type (e.g., Figure 1a) or reflected a cross-category insertion (e.g., Figure 1b), $\chi^2(1) = 11.53$, $p < .001$, whereas the response of 12 monolingual versus 3 bilingual children reflected a deletion, $\chi^2(1) = 7.71$, $p < .01$.

About half of the monolingual children but only one fourth of the bilinguals exhibited deletion as a significant change. This difference was reflected in the drawings' complexity scores. A comparison of the Kellogg Scale scores revealed that among the monolinguals, the regular flower was drawn with a higher level of complexity (Wilcoxon signed rank test, $Z = 2.12$, $p < .04$), whereas no such difference was observed for the bilinguals ($Z = 1.52$, $p > .1$).

Table 1
Language and Drawing Skills in Monolingual and Bilingual Children

	Study 1					Study 2				
	Monolinguals		Bilinguals			Monolinguals		Bilinguals		
	M	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
PPVT–IIIB (Hebrew)	58.16	13.48	59.48	16.51	0.31	73.60	25.09	50.13	14.72	3.14**
PPVT–IIIA			50.80	16.16				44.93	16.29	
Visual reception	12.83	4.11	12.64	4.31	0.16	14.63	5.92	15.53	6.09	0.18
Digit span	—	—	—	—		6.60	3.08	6.73	1.75	0.15
Drawing complexity scores										
Drawing summer	6.96	1.66	7.14	1.87	1.26	—	—	—	—	
Regular flower	6.86	2.62	7.58	2.54	0.78	7.67	2.06	8.36	1.60	1.00
Nonexistent flower	6.24	2.32	7.02	2.16	1.29	6.73	3.26	7.79	2.32	0.99
Regular house	—	—	—	—		7.67	2.67	8.07	1.97	0.52
Nonexistent house	—	—	—	—		7.73	1.83	7.21	2.51	0.63

Note. PPVT = Peabody Picture Vocabulary Test.
$^{**}p < .01$.

Table 2
Production of the Strange Flower Task of 4-Year-Old English–Hebrew Bilinguals and Hebrew Monolinguals

	Study 1			Study 2				
	Flower			Flower		House		
	Mono. $n = 25$	Bi. $n = 25$	χ^2	Mono. $n = 15$	Bi. $n = 15$	Mono. $n = 15$	Bi. $n = 14$	Z
Succeeded	16	21	2.60	11	13	9	12	1.49
Deletion	13	6	6.35*	5	4	5	5	0.48
Insertion	2	2	0.00	1	2	0	1	0.60
Element's shape	6	2	2.38	3	3	3	4	0.23
Whole shape	2	6	2.38	3	1	3	5	0.55
Orientation/position	2	2	0.00	0	0	2	2	0.00
Cross-category insertion	2	13	11.53***	2	7	3	8	2.96**

Note. In Study 2, the Wilcoxon test was used; Z statistics were computed for 29 of 30 children who had data on both drawings. Mono. = monolingual children; Bi. = bilingual children.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

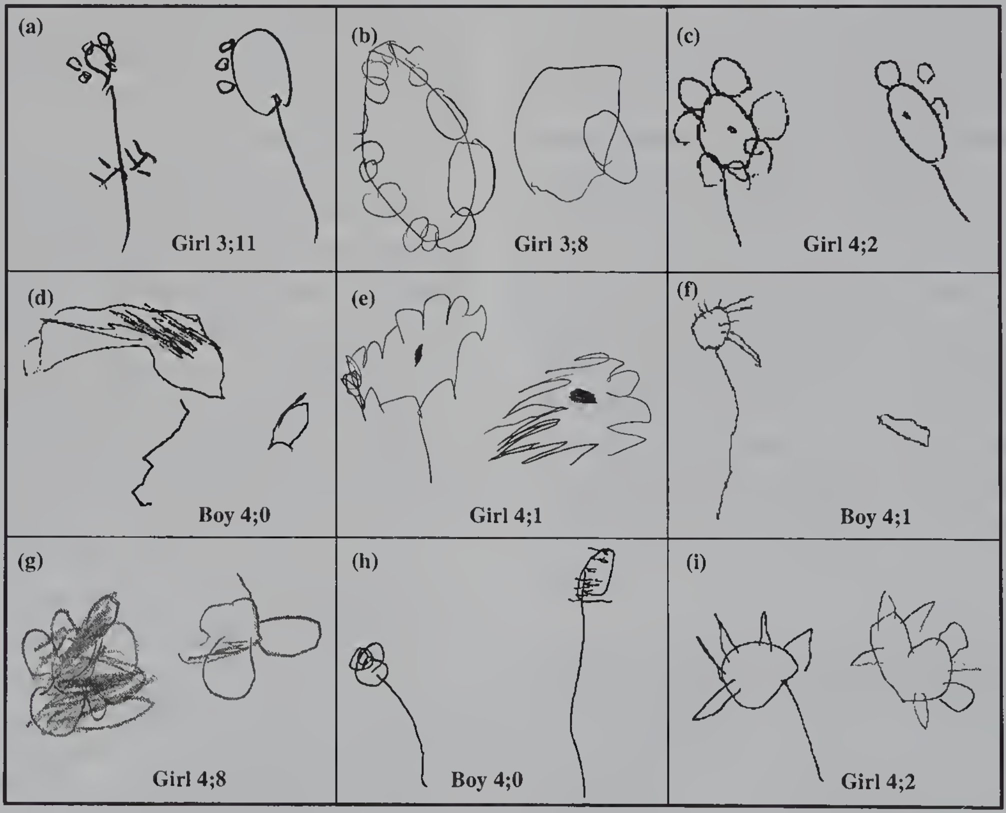


Figure 1. Deletion drawings by monolingual children.
Note. The child's verbal response and (in parentheses) the Kellogg Scale scores for regular and strange flowers are presented for each drawing. (a) "It's missing leaves" (9; 6). (b) "This is not real a flower because it has only one petal" (7; 4.5). (c) Didn't answer (9; 6). (d) "This is not real a flower because it doesn't have a stalk" (4.5; 4). (e) "This is not real a flower because it doesn't have roots" (9; 8). (f) "This is not real a flower; it doesn't have a stalk" (9; 8). (g) "This is not real a flower because it doesn't have the center" (8; 8). (h) "This is a flower in the wind, it doesn't have leaves" (8; 8). (i) "A flower which is a heart." Deletion (of stalk) is a subcategory in this example (9; 9).

The findings of Study 1 indicate that the 4-year-old English–Hebrew bilingual children and their monolingual Hebrew-speaking peers were comparable on a set of background measures. In spite of this similarity, the children produced different types of modifications: Monolingual children exhibited significantly more changes that involved deletion of elements within their drawing; bilingual children inserted significantly more features of other representational categories in their drawing of a flower. The results of Study 1 support the possibility that 4-year-old bilingual children display a

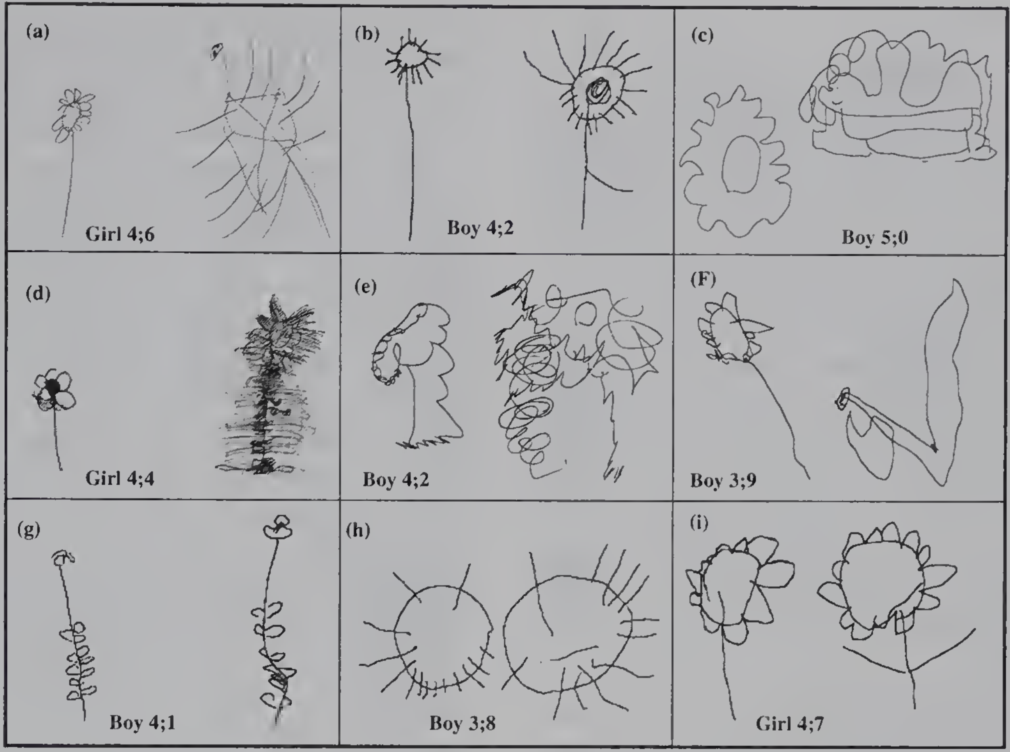


Figure 2. Cross-category insertion drawings of bilingual (English–Hebrew) children.
Note. The child's verbal response and (in parentheses) the Kellogg Scale scores for regular and strange flowers are presented for each drawing. Responses in italics were originally given in English. (a) *This is a giraffe flower* (9; 9.5). (b) *This is a flower with* a tail (9; 9). (c) *This is a camel flower* (8; 8.5). (d) *This is a lion flower with lots of hair. It has a lot of tails and shoes* (9; 10). (e) *This is a silly flower with arms and legs* (6; 4). (f) "This is not a real flower because it has arms and legs" (9; 6). (g) *This flower has teeth; it doesn't want to brush them*" (8.5; 9). (h) "*This is a flower* that knows how to smile" (6; 6). (i) "This is a flower with hands" (9; 9).

different pattern of changes than their monolingual peers, and that this difference is related to the larger portion of cross-category insertions seen in their drawings.

Hebrew and English differ in direction of writing: Hebrew is (read and) written from right to left. This may lead to greater graphic flexibility by English–Hebrew bilinguals. To control for this possibility, in Study 2 Arabic–Hebrew bilingual children were tested as a second bilingual group that share writing characteristics, including direction of writing, with their monolingual Hebrew-speaking peers.

Study 2

Method

Participants

Fifteen monolingual (6 boys and 9 girls) and 15 bilingual children (6 boys and 9 girls) participated in this study. The participants were 5-year-olds (range = 42–74 months, $M = 61.00$, $SD = 9.47$ and $M = 61.87$, $SD = 11.14$ months for the monolingual and bilingual children, respectively). The children participating in the study came from centrally located areas in Israel, came from middle-class families, and had parents who had higher education experiences. The children attended Hebrew-speaking kindergartens for at least 9 hr a day.

The bilingual children came from Muslim families. Bilingual participants attended a Hebrew-speaking kindergarten for not less than 2 years, at least from their 2nd year of life. These children came from an Arab town and attended a kindergarten in one of the nearby kibbutzim. Before their child entered kindergarten, the parents spoke with their children in Arabic. These children can therefore be classified as early sequential bilinguals (Kan & Kohnert, 2008). After their child entered kindergarten, at least one parent (usually the mother) spoke to the child in Hebrew. The parents' Hebrew level was sufficiently high to do so, partly because of their experiences in higher education, where the language of instruction is Hebrew, and experience at their work place.

Procedure

All children were observed individually in a quiet room at their kindergarten. Testing took place in 2–3 sessions (session time did not exceed 20 min). Arabic and Hebrew vocabulary tests were administered in separate sessions.

Measures

The PPVT–IIIA was translated into Arabic by an expert. The drawing summer assessment was not used (as a result of a ceiling effect), and a forward digit span test was added to ensure compatibility of groups.

Forward digit span. This test was used as an assessment of short-term memory. The children were presented with a random string of digits and were required to repeat the string back in the same order. The strings began with two digits and increased by one digit every two trials. One point was awarded for every correctly reproduced sequence, producing the possibility for two points for each string length. Testing ended when the child incorrectly reproduced both sequences at a given string length.

"Draw a flower/house that does not exist" test. The procedure detailed in Study 1 was used. The same phrasing was used to describe a flower and house that do not exist. The flower and house tasks were administered with 1–2 weeks separation, half of the children began with a house, and half with a flower. As a result of the absence of one child, only 14 bilingual children completed the house drawing. Cohen's κ coefficients for interrater agreement were above .92 ($p < .001$) in all change categories.

The drawing complexity scale. The complexity scoring procedure detailed in Study 1 was used except for Levels 8 and 9. Level 8 was applied when the house included either a door or window. Level 9 was applied when both appeared. Scores ranged between 2 and 10. Interjudge agreement was, on average, .94. Cohen's κ coefficients for interrater agreement on a categorical scale were, on average .76 ($p < .001$), and weighted κ coefficients for an ordinal scale were, on average .93 ($p < .001$). Kellogg Scale scores of drawing tasks were significantly correlated, $r(28) = .68$, $p < .001$, and significantly correlated with age $r(29,30) > .52$, $p < .001$.

Results and Discussion

Monolingual and bilingual children performed similarly on the forward digit span test, and had similar visual reception and drawing complexity scores (Table 1). However, monolinguals obtained significantly higher scores than their bilingual peers on the PPVT–III, $t(28)=3.14$, $p < .01$. A paired-sample *t* test indicated that the bilingual group had comparable language skills in Arabic and in Hebrew, $t(14) = 1.49$, *ns*. Other studies have shown that preschool bilingual children who first experience one language, but early on experience the second language, may have similar vocabularies in both languages (Kan & Kohnert, 2008).

Six change categories were used for both the flower and the house drawings, with a slight difference from Study 1: A change in the position or orientation of elements was exhibited only in nonexistent houses. Drawings were classified as having 0–4 category changes with most children (28/30 and 23/30 for the flower and house, respectively) producing 0–2 changes (Picard & Vinter, 2007). Overall, the children responded similarly to the nonexistent flower and house tasks: McNemar tests applied to each of the change categories revealed consistent modifications ($p > .13$); Wilcoxon tests showed that the overall number of change categories ($Z = .98$, *ns*), as well as drawing complexities ($Z = .99$, 1.06, *ns*, for the regular and nonexistent objects, respectively) were similar.

Table 2 presents the number of children in each of the change categories (Karmiloff-Smith, 1990). Compared with Study 1, children's modifications were distributed more evenly across the modification categories. Nevertheless, bilingual children produced significantly more cross-category insertions in the nonexistent flower as well as in the nonexistent house drawing (Figure 3). Eleven bilingual children produced at least one cross-category drawing compared with only four monolingual children, $\chi^2(1) = 6.53$, $p < .02$. In addition to human (5/15 drawings) and animal (3/15 drawings) features added to the flower and house, in Study 2 bilingual children replaced the flower (3/15 drawings) and house (4/15 drawings) with objects (e.g., a flower-door, a ball-house). Except for 2 children (1 monolingual and 1 bilingual) who did not provide a verbal response, children's verbal responses matched drawn modifications, and differences between the two groups in cross-category verbal responses paralleled differences in cross-category drawn modifications.

The findings of Study 2 indicate that the 5-year-old Arabic–Hebrew bilingual children and their monolingual Hebrew-speaking peers were comparable on a set of background measures but differed in their language skills. In spite of this difference, and similar to the findings of Study 1, bilinguals exhibited significantly more cross-category insertions in their drawings of nonexistent objects than their monolinguals peers. In Study 2, the children were requested to draw two nonexistent objects: a flower and a house. The results showed a consistent pattern of production and modifications in the

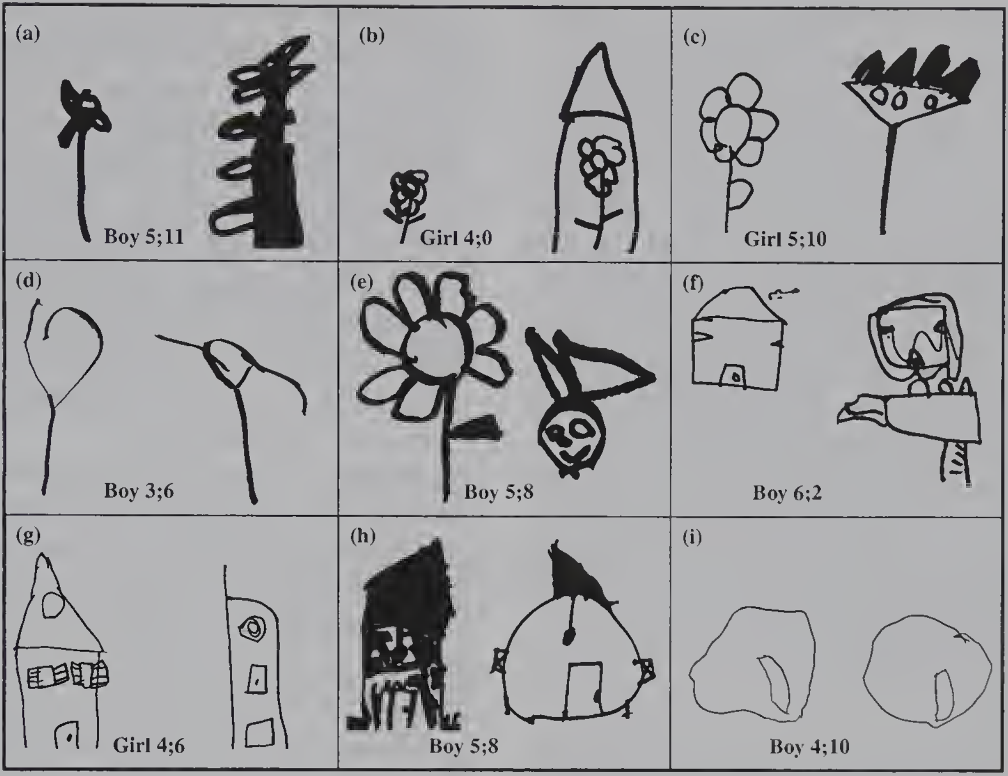


Figure 3. Cross-category insertion drawings of bilingual (Arabic–Hebrew) children.
Note. The child's verbal response and (in parentheses) the Kellogg Scale scores for regular and strange flowers are presented for each drawing. Responses in italics were originally given in Arabic. (a) "This is a tree flower" (9; 9). (b) *This flower is a door* (9; 9). (c) "A butterfly flower" (9; 8.5). (d) "This is a kite flower" (5; 6). (e) *This is a girl flower* (9; 9). (f) "A robot house" (9; 10). (g) "This is a chair house" (9; 8). (h) *This is a ball which is a house* (10; 9). (i) "This is an orange flower" (8; 8).

drawings of these objects. This may be related to the fact that both these objects are frequently produced by 5-year-olds. Furthermore, the rate of deletions of object features was similar across the groups, mainly because of a reduction in the rate of deletions produced by monolingual children. This reduction, albeit nonsignificant (Z = 1.93, in the comparison of two nonexistent flower drawing tasks made by monolinguals in Studies 1 and 2), may be related to the older age of the participants in Study 2.

General Discussion

Studies of representational flexibility consistently replicate the finding of a developmental constraint: Children less than 7 years do not respond to the task "draw an X that does not exist" with a complex intrarepresentational or interrepresentational innovation. The findings of this study indicate that 4- and 5-year-old bilingual children can produce a variety of interrepresentational modifications (i.e., cross-category insertions) in response to the experimenter's request to draw a flower (a house) that does not exist. These results underscore the difference between interrepresentational and intrarepresentational changes and suggest that bilinguals' language experiences may advance the former, but not on the latter type of representational flexibility.

Two samples of bilingual children participated in the study: 4-year-old English–Hebrew simultaneous bilingual children and 5-year-old Arabic–Hebrew early sequential bilingual children. As in previous studies, the atypical performance of the bilingual children was apparent already at age 4. Differences between groups were qualitative as well as quantitative and were probably not the result of the specific language groups compared, or specific object used in the task. Although the finding that 4- to 6-year-olds do not spontaneously produce cross-category insertions was replicated in several studies, and the advantage of 4-year-old as well as the 5-year-old bilinguals in introducing cross-category insertions was robust, future studies are needed to replicate the current results in a more heterogeneous sample of language backgrounds.

Why would bilinguals produce interrepresentational innovations earlier than their monolingual peers? Previous studies have shown that bilinguals exhibit better comprehension of figures or objects that comprise interrepresentational cues (Bialystok & Martin, 2004; Bialystok & Senman, 2004; Bialystok & Shapero, 2005). For example, bilinguals have demonstrated superior performance on an ambiguous figure task when required to identify the alternative image in a reversible figure (e.g., rat-man). It is possible that these children also have a superior ability to create (or at least imagine) figures that convey two representational meanings and that may enhance the frequency of drawings depicting interrepresentational flexibility. The results of several studies support the notion that bilinguals simultaneously keep their two language systems active and switch between them (e.g., Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008). Bilingual children spontaneously switch from one language to another in case of a lexical gap, creating mixed-language sentences (e.g., Lindholm & Padilla, 1978; Nicoladis & Secco, 2000). An early experience in mixing representational systems may as well enhance the frequency of interrepresentational modifications.

Several studies suggest that advantages in executive functions in bilinguals are specific rather than general (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). In line with these findings, the results of Studies 1 and 2 suggest that bilinguals did not demonstrate an overall flexibility advantage. Such an advantage would have been manifested if a higher rate of position and orientation changes, same-category insertions, and cross-category insertions would have been found in their drawings. Rather, bilinguals' performance was different from that of their monolingual peers in just one (of three) categories: Bilingual children produced more cross-category insertions.

Although drawing order was not systematically recorded, it was noticed that most bilingual children either added the cross-category item/s at the end of the drawing procedure (Figures 2b,d,g–i and 3d,f) or from first stroke made a new, blended, figure (Figure 2a,c,f and most drawings in Figure 3). For that to occur, children need to develop inhibitory control mechanisms allowing them to manage two conflicting demands (e.g., that the depicted figure would be recognizable *both* as a flower and as a giraffe). The production of position and orientation changes and same-category insertions typically involves graphic modifications that are executed in the middle of the drawing sequence (the child begins to draw, executes the graphic modification, and then goes on to finish the drawing as usual; Hollis & Low, 2005; Zhi et al., 1997). Graphic interruptions in the middle of the drawing procedure necessitate impulse control. Recent studies suggest that bilinguals have an advantage in the former inhibitory control mechanism (managing conflicting demands) but not in the latter (impulse control; Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). This may explain why bilinguals excelled only in the interrepresentational category of change. Drawing is a complex skill involving biomechanical, fine-motor, perceptual, cognitive, and other developmental competencies that do not vary by language (Braswell & Rosengren, 2008; Freeman & Adi-Japha, 2008; Morra, 2005; van Sommers, 1984). It is possible that these factors minimize the differences in representational flexibility that are expected in bilinguals.

Previous research has shown that bilingual children excel in cognitive flexibility tasks requiring a response to bivalent *perceptual* cues (e.g., Bialystok & Martin, 2004; Carlson & Meltzoff, 2008). This study extends the scope of these results to an imaginative cognitive flexibility task from a symbolic *production* domain (i.e., drawing), enabling an unlimited set of responses. Our findings suggest that the cognitive advantage of bilinguals is limited rather than general: Bilinguals produced more interrepresentational (but not intrarepresentational) modifications to their drawings of nonexistent object. The findings of this study provide further support for the notion that flexibility in the linguistic domain may facilitate flexibility in other nonlinguistic domains demonstrating the effects of experience (and in particular bilingualism) on children's development (Emmorey, Luk, Pyers, & Bialystok, 2008; Karmiloff-Smith, 2009). Representational as well as procedural aspects of graphomotor skills (e.g., drawing and writing, and in general any motor skill) improve with experience (e.g., Adi-Japha & Freeman, 2000, 2001; Adi-Japha, Karni, Parnes, Loewenschuss, & Vakil, 2008; Adi-Japha et al., 1998; Adi-Japha et al., 2007). Symbolic representation consists of other domains such as play or sculpture. Experience in symbolic domains is considered essential for cognitive development and the nurturing of executive skills (Diamond et al., 2007). Symbolic domains may provide bilingual children with a natural environment for presenting and enhancing their relative advantage.

References

Adi-Japha, E., & Freeman, N. H. (2000). Regulation and division of labor between cognitive systems controlling action. *Cognition*, *76*, 1–11.

Adi-Japha, E., & Freeman, N. H. (2001). Development of differentiation between writing and drawing systems. *Developmental Psychology*, *37*, 101–114.

Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., & Vakil, E. (2008). A shift in task routines during the learning of a motor skill: Group averaged data may mask critical phases in individuals' acquisition of skilled performance. *Journal of Experimental Psychology: Learning Memory and Cognition, 34*, 1544–1551.

Adi-Japha, E., Landau, Y., Frenkel, L., Teicher, M., Gross-Tzur, V., & Shalev, R. S. (2007). ADHD and dysgraphia: Underlying mechanisms. *Cortex, 43*, 700–709.

Adi-Japha, E., Levin, I., & Solomon, S. (1998). Emergence of representation in drawing: The relation between kinematic and referential aspects. *Cognitive Development, 13*, 23–49.

Arnheim, R. (1956). *Art and visual perception: A psychology of the creative eye*. London: Faber & Faber.

Berti, A. E., & Freeman, N. H. (1997). Representational change in resources for pictorial innovations: A three component analysis. *Cognitive Development, 12*, 501–522.

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24*, 560–567.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*, 636–644.

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified anti-saccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1341–1354.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*, 325–339.

Bialystok, E., & Senman, L. (2004). Executive processes in appearance–reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562–579.

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8*, 595–604.

Bialystok, E., Shenfield, T., & Codd, J. (2000). Languages, scripts, and the environment: Factors in developing concepts of print. *Developmental Psychology, 36*, 66–76.

Braswell, G., & Rosengren, K. (2008). The interaction of biomechanical and cognitive constraints in the production of children's drawing. In V. Lange-Küttner & A. Vinter (Eds.), *Drawing and the non-verbal mind* (pp. 123–138). Cambridge, UK: Cambridge University Press.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Psychology, 11*, 282–298.

Costa, A., Hernandez, M., & Sebastian-Galles, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59–86.

Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language, 24*, 3–24.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III*. Circle Pines, MN: American Guidance Service.

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science, 19*, 1201–1206.

Freeman, N. H. (1995). The emergence of a framework theory of pictorial reasoning. In C. Lange-Kuttner & G. V. Thomas (Eds.), *Drawing and looking: Theoretical approaches to pictorial representation in children* (pp. 135–146). London: Harvester.

Freeman, N. H., & Adi-Japha, E. (2008). Pictorial intention, action and interpretation. In V. Lange-Küttner & A. Vinter (Eds.), *Drawing and the non-verbal mind* (pp. 104–120). Cambridge, UK: Cambridge University Press.

Golomb, C. (1974). *Young children's sculpture and drawing*. Cambridge, MA: Harvard University.

Golomb, C. (1992). *The child's creation of a pictorial world*. Berkeley: University of California Press.

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish–English bilinguals: An fMRI study. *Neuroimage, 14*, 510–520.

Hollis, S., & Low, J. (2005). Karmiloff-Smith's RRM distinction between adjunctions and redescriptions: It's about time (and children's drawings). *British Journal of Developmental Psychology, 23*, 623–644.

Kan, P. F., & Kohnert, K. (2008). Fast mapping by bilingual preschool children. *Journal of Child Language, 35*, 495–514.

Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawing. *Cognition, 34*, 57–83.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology, 45*, 56–63.

Kellogg, R. (1970). *Analyzing children's art*. Palo Alto, CA: National Press Books.

Kirk, S. A., Mccarthy, J. J., & Kirk, W. (1968). *The Illinois Test of Psycholinguistic Abilities* (Rev. ed.). Urbana, IL: Institute for Research in Exceptional Children.

Lindholm, K., & Padilla, A. (1978). Language mixing in bilingual children. *Journal of Child Language, 5*, 327–335.

Luquet, G. H. (1927). *Le dessin enfantin* [Children's drawing]. Paris: Alcan.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*, 81–93.

Morra, S. (2005). Cognitive aspects of change in drawings: A non-Piagetian theoretical account. *British Journal of Developmental Psychology, 23*, 317–341.

Nicoladis, E., & Secco, G. (2000). Productive vocabulary and language choice. *First Language, 20*, 3–28.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space* (F. J. Langdon & J. L. Lunzer, Trans.). New York: Norton. (Original work published 1948)

Picard, D., & Vinter, A. (1999). Representational flexibility in children's drawings: Effects of age and verbal instructions. *British Journal of Developmental Psychology, 17*, 605–622.

Picard, D., & Vinter, A. (2006). Decomposing and connecting object representations in 5- to 9-year-old children's drawing behaviour. *British Journal of Developmental Psychology, 24*, 529–545.

Picard, D., & Vinter, A. (2007). Relationships between procedural rigidity and interrepresentational change in children's drawing behavior. *Child Development, 78*, 522–541.

Spensley, F., & Taylor, J. (1999). The development of cognitive flexibility: Evidence from children's drawings. *Human Development, 42*, 300–324.

Thomas, G. V., & Silk, A. M. J. (1989). *An introduction to the psychology of children's drawings*. New York: Harvester Wheatsheaf.

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*, 2706–2716.

van Sommers, P. (1984). *Drawing and cognition: Descriptive and experimental studies of graphic production processes*. Cambridge, UK: Cambridge University Press.

Zhi, Z., Thomas, G. V., & Robinson, E. J. (1997). Constraints on representational changes: Drawing a man with two heads. *British Journal of Developmental Psychology, 15*, 275–290.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1367–1375

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task

Sarah E. Berger
The College of Staten Island and the Graduate Center of the City University of New York

Karen E. Adolph and
Alisan E. Kavookjian
New York University

Using a means–means–ends problem-solving task, this study examined whether 16-month-old walking infants (N = 28) took into account the width of a bridge as a means for crossing a precipice and the location of a handrail as a means for augmenting balance on a narrow bridge. Infants were encouraged to cross from one platform to another over narrow and wide bridges located at various distances from a wooden handrail. Infants attempted to walk over the wide bridge more often than the narrow one and when the handrail was within reach. Infants demonstrated parallel problem solving by modifying exploratory behaviors and bridge-crossing strategies that simultaneously accounted for the spatial and functional relations between body and bridge, body and handrail, and bridge and handrail.

Means–ends problem solving is an important cognitive and motor achievement. It requires coordinating environmental supports with relevant body parts to extend the body's capabilities and monitoring multiple spatial and functional relations between actor, goal, and means. Pulling a cane to retrieve an out-of-reach toy, for example, requires recognition that the toy is beyond arm's reach and that a cane could extend the body's limits, implementation of an appropriate grasp of the handle, and understanding that the toy must be positioned inside the crook of the cane (e.g., Chen & Siegler, 2000; McCarty, Clifton, & Collard, 2001; van Leeuwen, Smitsman, & van Leeuwen, 1994). Changing frames of reference present a challenging "localization problem" (Lockman, 2000). Even preschoolers err in simple object-retrieval and spoon tasks after a change in the spatial relations between hook and toy or between hand and spoon (Smitsman & Cox, 2008). Task difficulty is compounded in *means–means*–ends tasks because the goal is so removed that multiple means must be employed to attain it (Santos, Rosati, Sproul, Spaulding, & Hauser, 2005). Multiple means may be implemented sequentially, as in using a short stick to rake in a longer stick, which could then be used to retrieve a distant toy, or in parallel, as in standing on a step stool while wielding a pole to reach a distant object (McGraw, 1935).

Typically, researchers study means–ends problem solving in infants by assessing their ability to use manual tools to retrieve an out-of-reach object and focusing on spatial relations between the actor, tool, and goal (e.g., Connolly & Dalgleish, 1989; Smitsman & Cox, 2008; Sommerville & Feldman, 2008; Steenbergen, van der Kamp, Smitsman, & Carson, 1997). A recent series of experiments took a different approach (Berger & Adolph, 2003; Berger, Adolph, & Lobo, 2005). The task involved the whole body, not just the hands, and we focused on means–means–ends problem solving rather than simple means–ends problem solving: To reach a goal at the other side of a deep precipice, infants had to recognize that a bridge could serve as a means for crossing and that a handrail could serve as a means for augmenting balance on narrow bridges. The spatial challenges of such bridge–

This research was supported by National Institute of Child Health and Human Development Grant R37-HD33486 to Karen E. Adolph and a CUNY Faculty Development Program Grant to Sarah E. Berger. Portions of this research were presented at the biennial meeting of the Cognitive Development Society, Park City, Utah, October 2003 and the International Conference on Infant Studies, Vancouver, British Columbia, March 2008. We gratefully acknowledge the members of the New York University Action Lab for their assistance with data collections and all of the infants and caregivers who participated in this research.

Correspondence concerning this article should be addressed to Sarah E. Berger, Department of Psychology, College of Staten Island and the Graduate Center, The City University of New York, 2800 Victory Boulevard, 4S-108, Staten Island, NY 10314. Electronic mail may be sent to sarah.berger@csi.cuny.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0004

handrail tasks require that the various body parts be coordinated with the location of the relevant environmental supports.

The bridge–handrail task requires a parallel means–means–ends solution. The body simultaneously links multiple elements—the feet balanced on the bridge and the hands grasping the handrail—as means for crossing the precipice. Such parallel coordination poses special problems because monitoring the top and bottom of the body simultaneously takes months to develop. Eleven-month-old cruising infants, for example, do not monitor their feet moving along the ground at the same time that they monitor their hands holding onto furniture for balance (Adolph, Berger, & Leo, in press). Although cruisers can see and feel their feet approaching a precipice, with their hands gripping a sturdy handrail, they step over the brink and fall. In contrast, 16-month-old walkers modify their crossing strategies on medium-sized bridges by slowing down, taking smaller steps, and turning sideways. To cross narrow bridges, they hang onto a sturdy handrail to augment balance. When bridges are impossibly narrow but no handrail is available, infants refuse to cross (Berger & Adolph, 2003). With only a wobbly handrail, they enhance balance control by pulling backward on the wobbly rail like a mountain-climbing rope (Berger et al., 2005).

Current Study

The current study focused on the complex spatial problems involved in the development of means–means–ends problem solving. We examined infants' ability to perceive the spatial relations among actor, goal, and two means in a new bridge–handrails task. We challenged 16-month-olds to cross bridges of varying widths located at various distances from a handrail. Infants had to monitor whether their bodies fit on the bridge, whether they could reach the handrail with their hands while their feet were on the bridge, and whether modification of handrail-use or bridge-crossing strategies would enable safe crossing. By varying bridge–handrail proximity, infants had to cope with the spatial relation between the two means for crossing the precipice. We expected greater success on wider bridges and when the handrail was in appropriate proximity to the bridge. We predicted that infants would devise strategies for crossing the bridges, but on narrow bridges far from the handrail, crossing would be impossible.

Method

Participants and Procedure

Twenty-eight infants (13 boys) were recruited through a commercial mailing list of families living in the New York City area. All were 16 months old (± 1 week) and born at term; most were White and middle class. In a structured interview (Adolph, Vereijken, & Shrout, 2003), parents reported the 1st day that infants walked independently 10 ft across a room. Walking experience ranged from 25 to 175 days ($M = 104.36$ days). Infants received a souvenir for participating.

Figure 1 shows the wooden bridge–handrail apparatus with starting and finishing platforms flanking a large precipice (Berger & Adolph, 2003; Berger et al., 2005). A 16- or 36-cm-wide bridge spanned the precipice and could be quickly changed between trials. A handrail ran parallel to the bridge. The bridge was placed at one of three locations relative to the handrail: flush, middle, and far, with the nearest edge of the bridge 0, 20.5, or 40 cm from the edge of the handrail, respectively. Previous work suggested that most infants would not require a handrail on the 36-cm bridge but would on the 16-cm bridge.

After infants were acclimated to the walkway, each bridge was presented twice at each location for a total of 12 trials. Six presentation orders were counterbalanced across sex. Infants began trials standing upright on the starting platform. Trials ended when infants crossed, fell, or remained on the starting platform for 45 s. Trials were videotaped with side and overhead views mixed onto a single video frame. After testing on the walkway, an experimenter measured infants' height, weight, and arm span. Sessions lasted approximately 60 min.

Data Coding

Data were coded from videotapes using MacSHAPA, a computerized video coding system (OpenSHAPA.org; Sanderson et al., 1994). A primary coder scored 100% of trials. To ensure interrater reliability, a secondary coder scored 25% of each infant's trials. For categorical and frequency measures, we calculated the proportion of trials that the two coders were in agreement. Interrater reliability ranged from .97 to 1.0; p values for all Cohen's Kappa coefficients < .01. For measures of duration, we calculated the correlation between coders' scores. Correlation coefficients ranged from $r = .98$ to $r = 1.0$. Only 2% of reliability trials

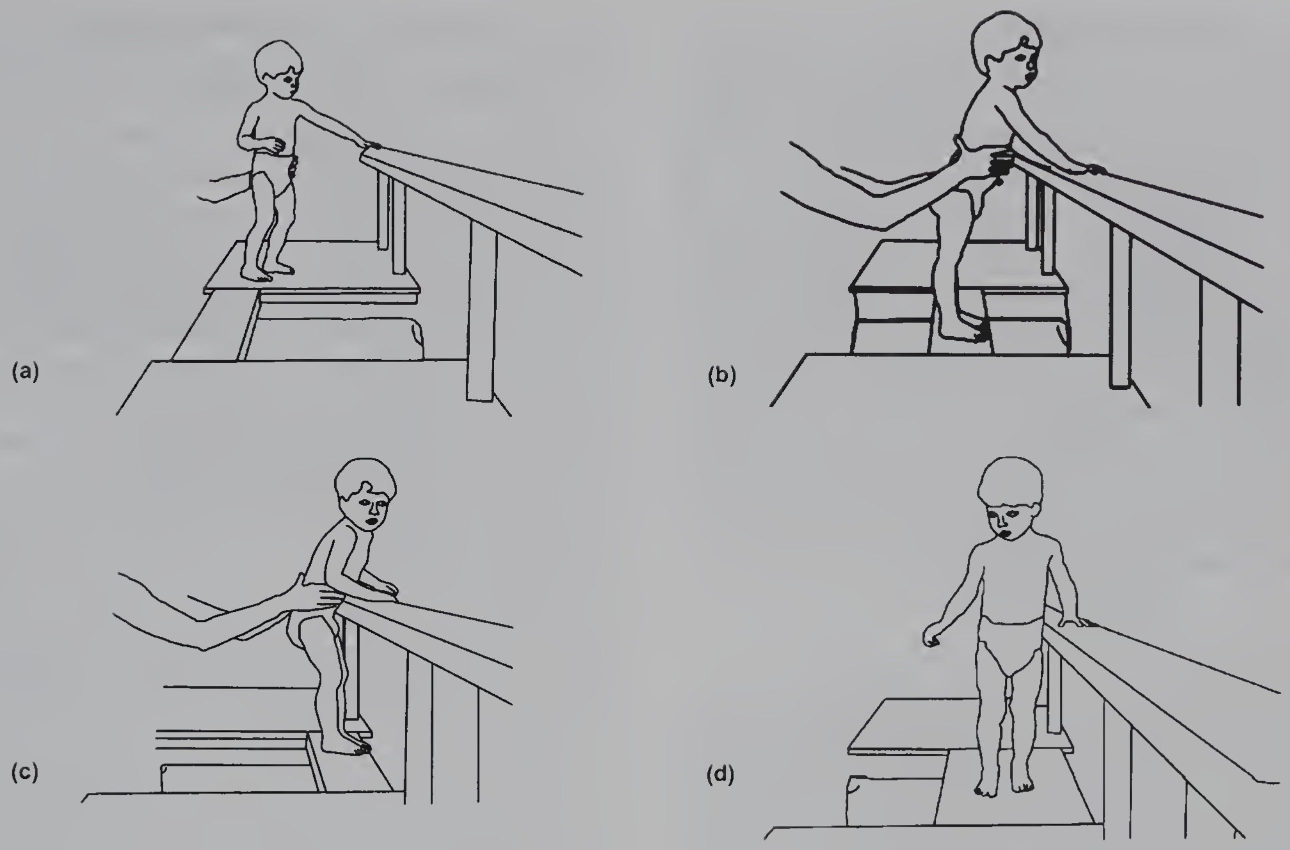


Figure 1. Walkway with varying bridge widths and handrail locations spanning a precipice (76.2 cm long × 76 cm wide × 74 cm deep) lined with 13.5 cm of foam padding along the sides and bottom. Apparatus is shown with (a–c) 16-cm and (d) 36-cm bridges at (a) far, (b) middle of the walkway, and (c,d) flush with the handrail locations.
Note. Caregivers (not shown) stood at the far end of the finishing platform (157 cm long × 76 cm wide × 86 cm high) and encouraged infants to cross by offering toys and cereal but did not draw infants' attention to the handrail or provide instructions. An experimenter (partially visible) followed alongside infants to ensure their safety. An assistant (not shown) adjusted the bridges and switched the handrails. To keep infants motivated, an easy, baseline 72-cm wide bridge was presented after each block of three trials.

resulted in discrepancies; these were resolved through discussion.

Crossing behaviors. Coders scored attempts to walk over the bridge (stepped onto the bridge with both feet) as successful if infants walked from one platform to the other without falling. Coders documented body position (facing forward, turned sideways) as infants stepped onto the bridge, whether they used the handrail (touched with ≥ 3 fingers for ≥ 0.5 s after stepping onto the bridge), and how they used the handrail (five combinations of body/arm positions). For successful attempts, coders measured number of steps to cross.

Exploration prior to embarking. Latency to step onto the bridge (from the moment that infants were released by the experimenter and faced the bridge or finishing platform until they stepped onto the bridge) included time spent in exploration and in displacement activities such as calling to the caregiver. Apparatus exploration included touching the handrail or bridge for ≥ 0.5 s before stepping onto the bridge. Exploration of the distance between bridge and handrail included shifting back and forth between bridge and handrail at least once in the same trial or gesturing toward the handrail by extending the arms while standing in front of the bridge.

Walking skill. To estimate infants' walking skill, average step length was calculated from videotape by dividing the distance between 2 points on the full walkway by the number of complete steps to walk from one point to the other.

Results

Attempts to cross, body position while crossing, and exploratory behaviors were analyzed with 2 (bridge width) × 3 (bridge location) repeated measures analyses of variance (ANOVAs). Main effects and interactions were followed by contrasts comparing each pair of locations for the two bridge widths and paired t tests comparing the two bridge widths at each of the three locations for a total of nine comparisons. A Bonferroni-adjusted alpha level corrected for experiment-wise error rates (overall $p = .05$).

Because so few infants attempted to cross the 16-cm bridge at the far location, we could not conduct repeated measures ANOVAs on infants' behaviors once they were on the bridge. Instead, we ran a series of planned comparisons for success rate, number of steps to cross the bridge, handrail use, and strategies for using the handrail.

Five children were missing either the 16-cm/middle or 36-cm/far width/location combination. These infants still received 12 trials total, but other trials were substituted due to experimenter error, for a total of 336 trials in the data set. Consequently, the number of participants included in repeated measures analyses was reduced from 28 to 23. The number of available trials was reduced for some measures because portions of trials were momentarily obscured from the camera view.

Crossing Behaviors

Infants attempted to walk over bridges on 72% of trials. An ANOVA on attempts to cross confirmed main effects for bridge width and bridge location (row A, Table 1; see Table 1 for all F and p values). Post hoc analyses revealed that infants attempted to cross the 36-cm bridge more often than the 16-cm bridge and attempts were less frequent when bridges were far from the handrail (all $ps \leq .03$; total height of bars in Figure 2a). Successful bridge crossings (striped regions in Figure 2a) were frequent on the 36-cm bridge at all locations but limited largely to the flush and middle locations on the 16-cm bridge (59% and 66% of attempts, respectively). Planned paired samples t tests on successful attempts revealed differences between the 16- and 36-cm bridges at all bridge locations (all $ps < .01$) and revealed that infants were more successful when the 16-cm bridge was at the middle than far location, $t(7) = 2.97$, $p < .03$.

Infants' modification of bridge-crossing and handrail-use strategies took into account the spatial relations between their bodies, the bridge, and the handrail. An ANOVA on proportion of trials that infants turned their bodies sideways to fit onto the bridge revealed main effects for bridge width and bridge location (row B, Table 1). Infants most frequently faced forward and put one foot in front of the other while crossing bridges (sparse striped and dotted regions of Figure 2a) but were likely to turn sideways, sliding one foot behind the other, when the bridge was narrow and when the bridge was in the flush or middle positions (dense striped and dotted regions of Figure 2a). At each location, infants walked forward on the 36-cm bridge more often than on the 16-cm bridge, all $ps < .05$. As with attempts and successes, infants also walked forward over the 16-cm bridge more often at the middle location than at the other locations (all $ps < .01$). We had assumed that the flush position would be most conducive for walking, and this was the arrangement in our previous studies. However, the narrow space of the flush position forced infants' bodies against the handrail and limited the number of effective whole-body solutions. The middle bridge location gave infants room to maneuver their bodies, leaving the handrail accessible.

On successful trials, infants took more steps to cross the 16-cm bridge than the 36-cm bridge at both middle and flush locations, all $ps < .01$ (Figure 2b). Only 3 infants successfully crossed the

Table 1
Main Effects and Interactions of Bridge Width and Handrail Location on Infants' Bridge-Crossing and Exploratory Behaviors

	Main effect width		Main effect location		Interaction width/location	
Outcome measure	*F*(*df*)	*p* value	*F*(*df*)	*p* value	*F*(*df*)	*p* value
Crossing behaviors						
A. Attempts	39.64(1, 22)	< .01	9.22(2, 44)	< .01	*ns*	
B. Body position	10.96(1, 22)	< .01	6.56(2, 44)	< .01	*ns*	
Exploratory behaviors						
C. Latency to step onto bridge	64.92(1, 20)	< .01	11.07(2, 40)	< .01	*ns*	
D. Touching handrail	31.04(1, 22)	< .01	22.44(2, 44)	< .01	5.92(2, 44)	< .01
E. Touching bridge	13.04(1, 22)	< .01	*ns*		*ns*	
Gauge distance between bridge and handrail						
F. Body shifts	27.39(1, 22)	< .01	15.74(2, 44)	< .01	9.70(2, 44)	< .01
G. Gestures toward handrail	14.72(1, 22)	< .01	*ns*		3.83(2, 44)	< .03

Note. Effect size for all main effects with $p < .01 = .53$. Effect size for main effect with $p < .03 = .47$.

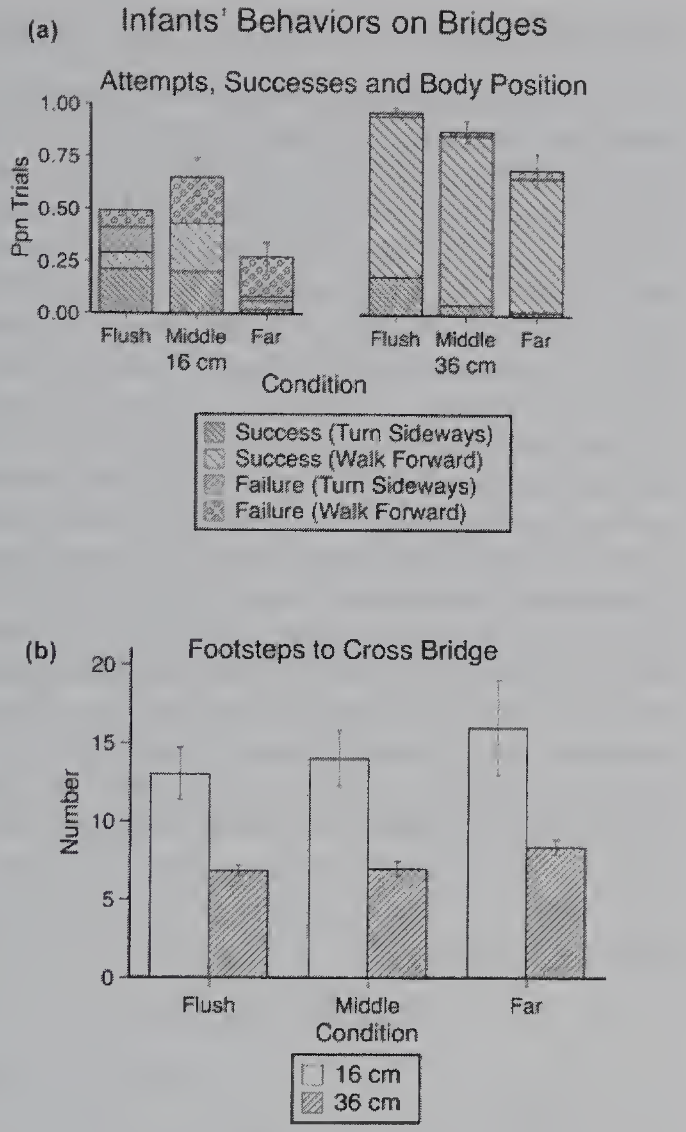


Figure 2. Effects of bridge width and bridge location on locomotor behaviors. (a) Proportion of trials infants attempted to walk, categorized by success and body position while crossing. (b) Number of footsteps infants took to walk over the bridge.
Note. Only 3 infants crossed the 16-cm bridge at the far location. Error bars denote standard errors.

16-cm bridge when it was far from the handrail, so analyses including that condition could not be performed. Location did not influence the number of steps to cross the 16-cm bridge. However, infants took more steps to cross the 36-cm bridge when it was far from the handrail than when it was flush with the handrail or in the middle of the walkway, $t(17) = 2.72$, $p = .01$ and $t(17) = 2.55$, $p = .02$, respectively.

Twenty-six infants used the handrail at least once. On the 90 trials when infants used the handrail, they held the handrail as they embarked or immediately after (within 2 s) stepping onto the bridge 65.6% of the time, suggesting that they coordinated both means in parallel prior to crossing. Planned paired-samples *t* tests confirmed that on attempted trials infants used the handrail more often on the 16-cm bridge than the 36-cm bridge when bridges were in the middle of the walkway, $t(17) = 10.31$, $p < .01$ (Figure 3). On the 16-cm bridge, infants never used the handrail at the far location. Handrail use at flush and middle locations did not differ. On the 36-cm bridge, only 1 child on one trial used the handrail at the far location. Surprisingly, infants used the handrail more often at the flush than middle and far locations, $t(26) = 8.37$, $p < .01$, and $t(18) = 7.51$, $p < .01$, respectively, even though touching was unnecessary (infants successfully crossed the 36-cm bridge without handrail use at the middle and far positions and in previous work; Berger & Adolph, 2003). Flush touches were cursory taps with one hand, as if noting rather than needing the handrail. When infants needed the handrail as a means for keeping balance, their touches were extended and they leaned toward the rail with their whole body.

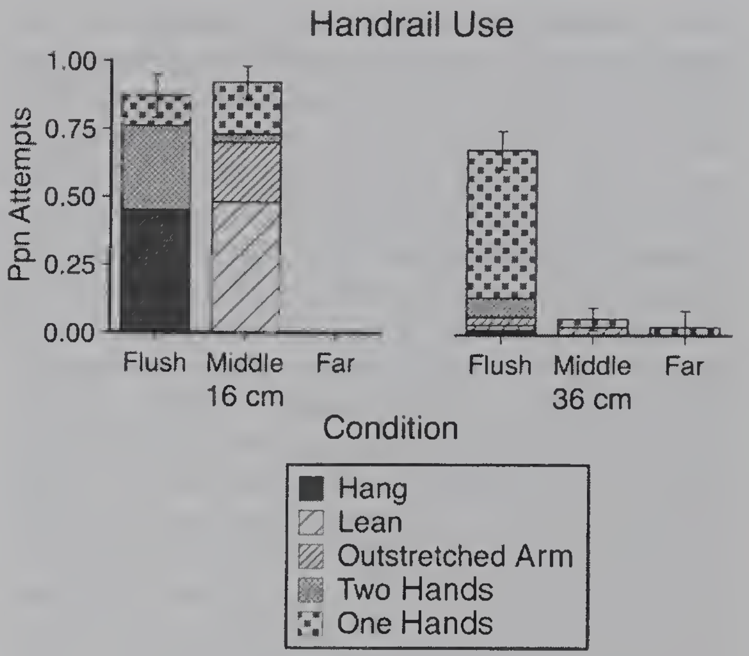


Figure 3. Proportion of attempted trials that infants used the handrail and their strategies for using the handrail at each bridge width/location combination.
Note. Infants could execute more than one type of strategy per trial but did not receive credit for performing the same strategy more than once per trial. Error bars denote standard errors.

Rates of handrail use were similar for the 16-cm flush, 16-cm middle, and 36-cm flush bridges, but infants displayed five qualitatively different strategies for coordinating placement of their hands on the handrail and their feet on the bridge, depending on bridge width and location. On half of the 16-cm flush trials, infants fit on the bridge by pressing the handrail against their chest and draping their bodies and arms over the railing (solid region in Figure 3 and illustrated in Figure 1c). On the other half, they used both hands to hold on as they walked sideways facing the handrail with their arms bent (dense checked region in Figure 3). On the 16-cm bridge in the middle of the walkway, infants turned their bodies sideways, leaned forward at the waist, and grabbed the handrail with two outstretched hands (sparse striped region in Figure 3 and illustrated in Figure 1b) or they walked forward with an arm fully outstretched to the side (dense striped region in Figure 3). In contrast, on the 36-cm bridge at the flush location, infants lightly touched the handrail with one hand as they walked (Figure 3 sparse checked region and illustrated in Figure 1d). The duration of infants' touches on the 16-cm flush (2.7 s) and 16-cm middle (2.85 s) bridges lasted over twice as long as they did on the 36-cm flush bridge (1.29 s), indicating that infants clung to the handrail on the 16-cm bridge but quickly tapped the 36-cm bridge as they passed. Thus, infants invented strategies for surmounting the distance between the bridge and handrail that were tailored to the width of the bridge and the distance from the handrail.

Exploration of Spatial Configuration Prior to Embarking

Infants engaged in a variety of information-gathering behaviors before deciding whether to cross the bridge. The ANOVA on latency to step onto the bridge confirmed main effects for bridge width and bridge location (row C, Table 1). Post hoc comparisons revealed longer latencies on the 16-cm bridge than on the 36-cm bridge and longer latencies at the far location than at the flush or middle locations (all *ps* < .03; Figure 4a).

Exploratory touching of the handrail and bridge decreased on the wider bridge and as the bridge moved farther from the handrail (rows D and E, Table 1; Figure 4b). The ANOVA for handrail exploration confirmed main effects for bridge width, bridge location, and an interaction between the two factors; the ANOVA for bridge exploration showed only a main effect for location. Post hoc comparisons revealed more exploration of the handrails on the 16-cm bridge than the 36-cm bridge at all locations (all *ps* < .04) and more bridge

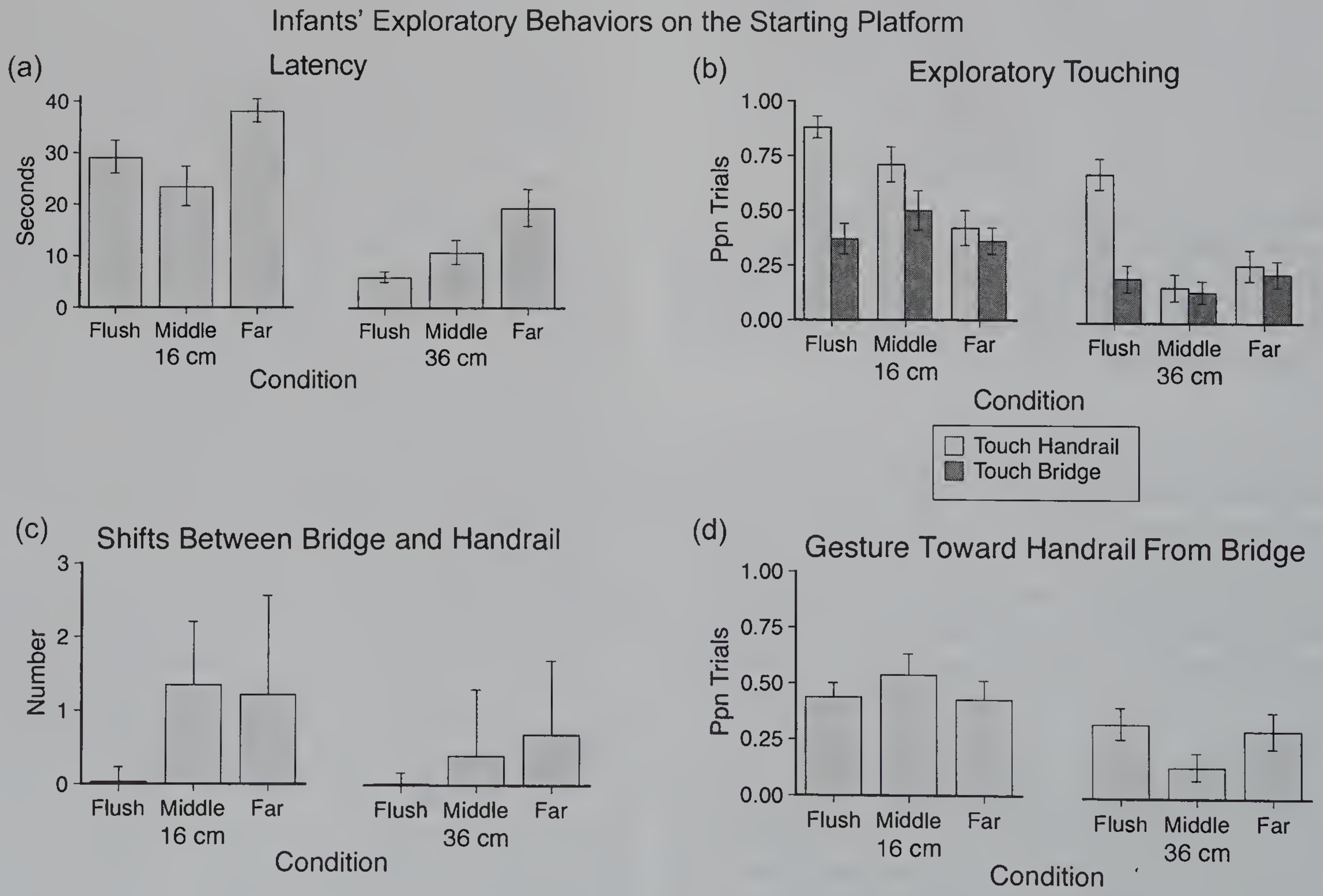


Figure 4. Effects of bridge width and bridge location relative to the handrail on exploratory behaviors. (a) Latencies to step onto the bridge. (b) Touching the handrail (light bars) and bridge (dark bars). Exploration of the distance between bridge and handrail by (c) shifting back and forth between the bridge and the handrail and (d) gesturing toward the handrail while standing at the bridge.
Note. Error bars denote standard errors.

exploration on the 16-cm middle location only, $t(22) = 3.85$, $p < .01$.

To gauge the distance between the bridge and handrail and decide whether and how their bodies could span the distance between the two means, infants shifted their bodies back and forth between the bridge and handrail while still on the starting platform (Figure 4c) and leaned their bodies or stretched out their arms toward the handrail when they were standing in front of the bridge (Figure 4d). ANOVAs on frequency of body shifts and gestures toward the handrail revealed main effects for bridge width and interactions between bridge width and bridge location (rows F and G, Table 1); the ANOVA on exploratory shifts also revealed a main effect for bridge location. Post hoc comparisons revealed that infants vacillated between the bridge and handrail more often on the 16-cm bridge than on the 36-cm bridge at both the middle and far locations, shifted more often at the middle and far locations than at the flush location on both bridge widths (all ps $< .03$), and gestured toward the handrail more often when the bridge was 16 cm than when it was 36 cm at the middle location, $t(22) = 4.03$, $p < .01$.

Individual Differences

On average, infants were 78.51 cm tall (range = 70.05–91.00 cm) and their mean arm span from fingertip to fingertip with arms outstretched was 81.69 cm (range = 76.20–88.45 cm). This means that the handrail fell at approximately elbow level and, in the far position, the handrail was beyond their reach unless they leaned nearly parallel to the ground. Infants' mean weight was 10.42 kg (range = 8.54–12.75 kg). Taller infants walked faster on baseline trials, $r(21) = -.51$, $p < .02$. Neither body dimensions nor days since walking onset were correlated with exploratory behaviors, attempts to cross the bridges, or success at crossing. However, infants' arm span was significantly correlated with handrail use. The longer their arms, the more likely they were to use the handrail, $r(23) = .42$, $p < .05$, and to carry out strategies for spanning the distance between the handrail and the bridge (leaning their bodies as they walked or walking with an arm outstretched to reach the handrail), $r(23) = .45$, $p = .03$.

Walking proficiency was related to performance on the bridges. The longer the distance between each footstep on baseline trials (an indication of mature gait), the shorter infants' latencies to step onto the bridge, $r(28) = -.46$, $p < .02$, the less likely they were to explore the distance between bridge and handrail, $r(28) = -.44$, $p < .02$, and the more likely they were to cross successfully, $r(28) = .47$, $p < .02$.

Handrail use and success were correlated with strategies for coping with the distance between the bridge and handrail. The more often infants gestured toward the handrail or shifted back and forth between the bridge and handrail on the starting platform, the more likely they were to use the handrail once on the bridge (Figure 1a; top panel Table 2). The more often infants leaned their bodies, stretched out their arms so that they could reach the handrail, turned their bodies sideways, or used both hands to hold onto the handrail when they were on the bridge, the more likely they were to cross successfully (bottom panel Table 2). For trials on the 16-cm bridge only, success was also related to whether infants used the handrail, $r(28) = .60$, $p < .01$.

Discussion

Means–Means–Ends Problem Solving

We examined infants' solutions to the localization problem of coordinating "multiple and mobile frames of reference" (Lockman, 2000, p. 140) in a whole-body, means–means–ends task. This work advanced previous studies of infants' problem solving in three ways. First, infants had to

Table 2

Correlations Between Infants' Exploration of Distance Strategies and Handrail Use After Embarking and Crossing Strategies and Success

Strategy type	Outcome measure	
	Use handrail	
Exploration of distance	*r(df)*	*p* value
Lean body or stretch arms toward handrail	.42(28)	< .03
Shift body between bridge and handrail	.41(28)	< .03
	Success	
Crossing strategy	*r(df)*	*p* value
Lean body	.55(28)	< .01
Walk with outstretched arms	.58(28)	< .01
Turn sideways on bridge	.71(28)	< .01
Use both hands to hold handrail	.69(28)	< .01

coordinate three key spatial relations on each trial—bridge and handrail, bridge and body, and handrail and body—rather than coordinate two frames of reference, such as object and surface or actor and tool (e.g., Bourgeois, Khawar, Neal, & Lockman, 2005). Second, infants had to monitor changes in spatial relations of the two means due to variations in bridge location. Third, with each step on the bridge, infants had to update their assessment of the spatial relations.

Sixteen-month-olds were active problem solvers. When infants attempted to cross narrow bridges, they juggled fitting their bodies onto the bridge, testing whether they needed a handrail to augment their balance, and orienting their bodies so that their upper extremities were securely on the handrail while their feet were firmly on the bridge. Infants viewed the bridge as a means for crossing the walkway, the handrail as a means for augmenting balance, and their own bodies as the link between the two. Failures were infrequent, suggesting that infants coordinated three simultaneously changing spatial relations. This finding is the first systematic demonstration of parallel coordination of multiple means in a problem-solving task in human infants.

Bridging the Gap: Understanding Spatial Relations

This experiment bridged the gap between research on spatial reasoning and studies of locomotor development. Typically, means–means–ends tasks require sequential implementation of means, such as using a knife to slice a lemon and then a reamer to squeeze out the juice. Alternatively, multiple means can be carried out in parallel, such as using salad servers to dole out a serving. Logically, infants would have to perceive a bridge as too narrow before considering whether a handrail was a viable means for augmenting balance. On wide bridges, infants ignore the handrail; they run straight across without touching the handrail (Berger & Adolph, 2003). However, once infants recognized a bridge as too narrow to serve as a sole means for crossing the precipice, they considered the spatial relations of the two means relative to each other and to themselves. Handrail and bridge-crossing strategies were highly sensitive to the spatial configuration of each condition. On most narrow bridge crossings, infants were already holding the rail as they stepped onto the bridge or grabbed onto it immediately after, indicating consideration of the two means in parallel, rather than in sequence, before going onto the bridge.

Previously (Berger et al., 2005), we postulated that implementing a means might lag behind recognizing a means because of a reported delay between the time toddlers choose the appropriate tool to solve a problem at about 21 months old and the time they actually use the tool to solve the problem at about 31 months old (Chen & Siegler, 2000). But this might not be true. Instead, success may result from the opportunity to devise clever, new solutions. Infants used their bodies to assess distance with the explicit purpose of determining whether the handrail and bridge in combination could be used as means to solve the problem. Infants devised new ways of using the handrail unobserved in previous bridge and handrail studies such as leaning forward while walking sideways (Figure 1b). Forcing infants to choose from a fixed set of behaviors determined by the experimenter and to use tools arranged inches from midline (e.g., Chen & Siegler, 2000; Smitsman & Cox, 2008) may artificially limit infants' ability to figure out the means to an end. In real life, tools may be at a great distance, and creativity is crucial for devising solutions to novel problems.

The suite of bridge–handrail studies showcases impressive problem solving in 16-month-old infants. Infants demonstrated knowledge about needing a suitable floor for crossing, about recruiting implements from the environment to enhance balance control, and about modifying typical locomotor strategies in response to ongoing challenges. The task for future research is to trace the developmental history beginning with cruising infants' first recognition that they can use a handrail as a means to support themselves to the sophisticated behaviors demonstrated by 16-month-olds as they solve means–means–ends problems.

References

Adolph, K. E., Berger, S. E., & Leo, A. (in press). Developmental continuity? Crawling, cruising, and walking. *Developmental Science*.

Adolph, K. E., Vereijken, B., & Shrout, P. E. (2003). What changes in infant walking and why. *Child Development*, *74*, 475–497.

Berger, S. E., & Adolph, K. E. (2003). Infants use handrails as tools in a locomotor task. *Developmental Psychology*, *39*, 594–605.

Berger, S. E., Adolph, K. E., & Lobo, S. A. (2005). Out of the toolbox: Toddlers differentiate wobbly and wooden handrails. *Child Development*, *76*, 1294–1307.

Bourgeois, K. S., Khawar, A. W., Neal, S. A., & Lockman, J. J. (2005). Infant manual exploration of objects, surfaces and their interrelations. *Infancy*, *8*, 233–252.

Chen, Z., & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development*, *65*(2, Serial No. 261).

Connolly, K., & Dalgleish, M. (1989). The emergence of a tool-using skill in infancy. *Developmental Psychology*, *25*, 894–912.

Lockman, J. J. (2000). A perception-action perspective on tool use development. *Child Development*, *71*, 137–144.

McCarty, M. E., Clifton, R. K., & Collard, R. (2001). The beginnings of tool use by infants and toddlers. *Infancy*, *2*, 885–893.

McGraw, M. B. (1935). *Growth: A study of Johnny and Jimmy*. New York: Appleton-Century.

Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Wantanabe, L. M., & James, J. M. (1994). MacShapa and the enterprise of Exploratory Sequential Data (ESDA). *International Journal of Human Computer Studies*, *41*, 633–681.

Santos, L. R., Rosati, A., Sproul, C., Spaulding, B., & Hauser, M. D. (2005). Means-means-end tool choice in cotton-top tamarins (*Saguinus oedipus*): Finding the limits on primates' knowledge of tools. *Animal Cognition*, *8*, 236–246.

Smitsman, A. W., & Cox, R. F. A. (2008). Perseveration in tool use: A window for understanding the dynamics of the action-selection process. *Infancy*, *13*, 249–269.

Sommerville, J., & Feldman, E. (2008, March). *Tool use in the 1st year of life: Infants select appropriate tool use actions based on tool and goal object properties*. Paper presented at the biennial international conference on Infant Studies, Vancouver, Canada.

Steenbergen, B., van der Kamp, J., Smitsman, A. W., & Carson, R. G. (1997). Spoon handling in two- to four-year-old children. *Ecological Psychology*, *9*, 113–129.

van Leeuwen, L., Smitsman, A. W., & van Leeuwen, C. (1994). Affordances, perceptual complexity, and the development of tool use. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 174–191.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1376–1383

What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning

Christopher T. Fennell
University of Ottawa

Sandra R. Waxman
Northwestern University

Past research has uncovered a surprising paradox: Although 14-month-olds have exquisite phonetic discrimination skills (e.g., distinguishing [b] from [d]), they have difficulty using phonetic detail when mapping *novel* words to objects in laboratory tasks (confusing *bin* and *din*). While some have attributed infants' difficulty to immature word learning abilities, the hypothesis presented herein is that infants are powerful word learners and this apparent difficulty occurs only when the referential status of the novel word is unclear. Across 2 experiments, 14-month-old infants ($N = 44$) used phonetic detail to map novel words to objects when conditions were conducive to word-referent mapping (clear sentential contexts and word-referent training), thus revealing no fundamental discontinuity in its use from speech perception to word learning.

In their 1st year of life, infants become attuned to the speech sounds of their native language. Infants' initial sensitivity to a universal set of phonetic categories is shaped by experience with the ambient language. By their first birthdays, they refine their sensitivity to those phonetic categories that carry meaning in their native language (Kuhl et al., 2006; Werker & Tees, 1984). During this same developmental period, infants become increasingly proficient word learners. By 12 months, they have acquired dozens of words and produce at least a handful on their own (Fenson et al., 1994). By 14 months, infants begin to use not only the word itself but also the linguistic context in which it is presented as a cue to meaning. They identify novel words presented as count nouns (e.g., ''This is a toma'') and map them specifically to objects and object categories (Waxman & Booth, 2001).

Infants' early advances in speech perception and word learning are impressive, but recent investigations have uncovered a surprising paradox: Fourteen-month-old infants appear unable to effectively recruit their fine sensitivity to native language phonetic detail when mapping novel words to meaning (Pater, Stager, & Werker, 2004; Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). Our goal in this article is to demonstrate that this paradox is apparent only: Infants successfully recruit phonetic detail in novel word–object mappings when it is clear that the word is intended to refer to the object.

To begin, we review the evidence for this apparent paradox. We focus primarily on 14-month-olds' performance in the Switch task (Werker, Cohen, Lloyd, Casasola, & Stager, 1998) because it is in this task that their failure to recruit phonetic detail in establishing word–object mappings has been revealed. Our review underscores the precision of the now-classic Switch task and its value in developmental research, but also brings to light an important caveat: There is nothing in the design of this task that makes it clear that the novel word involves reference.

In the Switch task, infants are first introduced to a novel word–object pairing. They hear a novel word, presented repeatedly, as they observe an image of a brightly colored novel object moving back and forth across a screen. This pairing continues until infants find it familiar, as indicated by a criterial decline in looking. In the subsequent test phase, infants witness two types of test trials, presented sequentially. On ''same'' trials, the original word–object pairing is preserved. On ''switch''

We thank J. Werker and E. Leddon for discussions, A. Weisleder and L. Plotkin for testing, and the participating parents and infants. NIH R01-HD30410 (S.W.) and a SSHRC Postdoctoral Fellowship (C.F.) provided research support.

Correspondence concerning this article should be addressed to Christopher T. Fennell, School of Psychology and Department of Linguistics, University of Ottawa, Ottawa, ON, Canada K1N 6N5, or Sandra R. Waxman, Department of Psychology, Northwestern University, Evanston, IL 60208. Electronic mail may be sent to fennell@uottawa.ca or swaxman@northwestern.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0005

trials, the pairing is disrupted: The object is paired with a different word. The logic is straightforward: If infants detect this disruption, they should look longer on "switch" than on "same" trials. However, when the two novel words are phonetically similar (e.g., *bin* vs. *din*), 14-month-olds fail to detect the mismatch (e.g., Stager & Werker, 1997).

Considerable attention has been devoted to identifying the source of 14-month-olds' surprising failure. It does not stem from a frank inability to connect novel words with novel objects in the Switch task. When the novel words are completely phonetically dissimilar (*lif* vs. *neem*), 14-month-olds detect the disruption readily. Neither does their failure stem from an inability to perceive the close contrast (e.g., *b* vs. *d*). When the very same contrast is presented within the context of a standard speech perception task, 14-month-olds' sensitivity is robust (Stager & Werker, 1997). If 14-month-olds can master the demands of the switch task and can also detect the relevant contrast, what accounts for their surprising difficulty? Some have proposed that infants' difficulty reflects a discontinuity between the perceptually-based *phonetic* representations that infants refine over the first year, and the more abstract *phonological* representations that are required to support word learning (Brown & Matthews, 1997). Recent investigations cast doubt on this interpretation, however, because infants as young as 11 months are sensitive to fine phonetic distinctions between *known* words (e.g., *ball* vs. *doll*; Swingley, 2005). This finding, coupled with the fact that infants successfully incorporate detailed acoustic-phonetic information in their representations of novel as well as familiar words (Fisher, Church, & Chambers, 2004) indicates that in principle, infants can bring their appreciation of phonetic detail to bear in word learning.

Why, then, do 14-month-olds fail to recruit their finely tuned phonetic sensitivities when mapping words to objects in the Switch task? This question is especially compelling in face of evidence that infants this age are spontaneously adding words to their lexicons in earnest (Fenson et al., 1994) and even use unstressed grammatical elements (determiners) as cues to meaning (e.g., Waxman, 1999). Thus far, efforts to address this paradoxical finding in the Switch task have appealed either to general difficulties associated with word learning (Werker & Fennell, 2004) or to a host of perceptual, attentional and memory factors (e.g., Rost & McMurray, 2009; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009).

We offer a different view. We propose that infants do indeed harness their phonetic sensitivities in word learning, but that their apparent failures to do so in previous work reflect the ambiguous referential status of the novel words. To appreciate this ambiguous status, notice that in all previous implementations of the Switch task, novel words were presented in isolation. While this ensured that the speech stimuli would be identical to those presented in standard speech perception tasks, presenting words in isolation can have adverse consequences with respect to word learning. First, infants process isolated words *less* efficiently than words presented within phrases (Fernald & Hurtado, 2006). Second, words rarely occur in isolation in naturally occurring speech (Aslin, Woodward, LaMendola, & Bever, 1996; Brent & Siskind, 2001). Finally, and perhaps most importantly for the purposes of establishing meaning, when words do appear in isolation in naturally occurring speech, they are rarely names for objects. Instead, isolated words tend to be proper names ("Daddy!"), commands ("Stop!") or exclamations ("Wow!"). Importantly, infants are sensitive to this state of affairs. When novel words are presented within naming phrases ("Look at the blick!"), infants readily map them to the objects with which they occur. But when the very same words are presented in isolation ("Look! Blick!"), infants fail to establish a word–object mapping (Fulkerson & Waxman, 2007; Namy & Waxman, 2000).

If 14-month-old infants' failure to use phonetic detail in the Switch task is indeed a consequence of the ambiguous referential status of words, then when referential status is clarified, infants should successfully recruit their finely tuned phonetic sensitivities to establish a word–object mapping. We designed two experiments to test this proposal. In both, we introduced 14-month-olds to the same words, objects, and methods used in previous work (Pater et al., 2004; Stager & Werker, 1997; Werker et al., 2002), but systematically manipulated the manner in which we introduced the novel words in order to increase or decrease the clarity of the word-referent mapping.

Experiment 1

In this experiment, we embedded novel words in typical naming phrases. Our choice of naming phrases was guided by evidence that 14-month-olds interpret novel words presented as count nouns as names for objects and object categories (Booth & Waxman, 2009; Waxman & Booth, 2001).

Method

Participants

Sixteen English-learning (> 80% exposure) healthy full-term infants from greater Chicago participated (M = 14.7 months, range = 13.5–15.5 months), primarily from Caucasian middle- to upper-middle-class families recruited via mailings. An additional 11 infants were excluded due to: restlessness (10) or parental interference (1).

Stimuli

The stimuli were nearly identical to those in previous work, except that the novel words were presented in naming phrases.

Auditory. We recorded an English-speaking female producing three novel CVC words (*bin*, *din*, *neem*) in infant-direct speech within the context of seven different introductory phrases: Look. It's the _____. There's the ______. Do you see the _____? Look at the _____. Where's the ____? Here's the ______. I like the _____. We ensured that infants could not rely upon subtle sound differences in the introductory phrases to discriminate the target words. We spliced the introductory phrases from the *neem* stimuli just before the determiner (e.g., "*Look at* | the neem."), copied them, and spliced the determiners and nouns from the *bin* and *din* sentences onto the copies. This resulted in sentences where the introductory phrases were identical, as indicated in italics: "*Look at* the neem!" "*Look at* the bin!" and "*Look at* the din!" One isolated exemplar of the novel word occurred at the beginning of each trial so that infants would hear the relevant phonetic information first; another exemplar occurred at the end. *Neem* phrases were used for pre- and posttests; *bin* and *din* phrases were used for habituation and test trials.

Visual. See Figure 1. The habituation object moved horizontally across the screen. The pre- and posttest object remained stationary, while its wheel rotated.

Procedure

See Figure 1. After parent(s) completed the MacArthur Bates Communicative Development Inventory (MCDI), the infant and one caregiver were escorted to a quiet room (3.12 × 3.89 m), lit by a shaded 60-Watt lamp situated to the left and slightly behind the infant. Visual stimuli were projected onto a screen surrounded by black cloth 1.93 m from the infant. Audio stimuli were delivered at 65 dB, ± 5 dB, over two speakers located below the screen. As a masking device, caregivers listened to music over headphones. Infants' looking times were monitored via a hidden digital videocamera. The experiment was administered with the Habit X program (Cohen, Atkinson, & Chaput, 2004). The experimenter, who wore headphones throughout the procedure, was blind to trial type. Each 20-s trial began when the infant fixated on an onscreen attention getter. To begin, infants were presented with a pretest trial, involving a waterwheel object and the *neem* phrases. The habituation phase commenced immediately thereafter and ended when the infant's looking time across a two-trial block decreased to 65% of the longest block, or after 24 habituation trials (mean length = 10.55 trials, 95 tokens; SD = 6.22, 55.98), after which the test phase began. The word used in habituation phrases (*bin*, *din*) and the presentation order of the test trials were counterbalanced. A posttest trial,

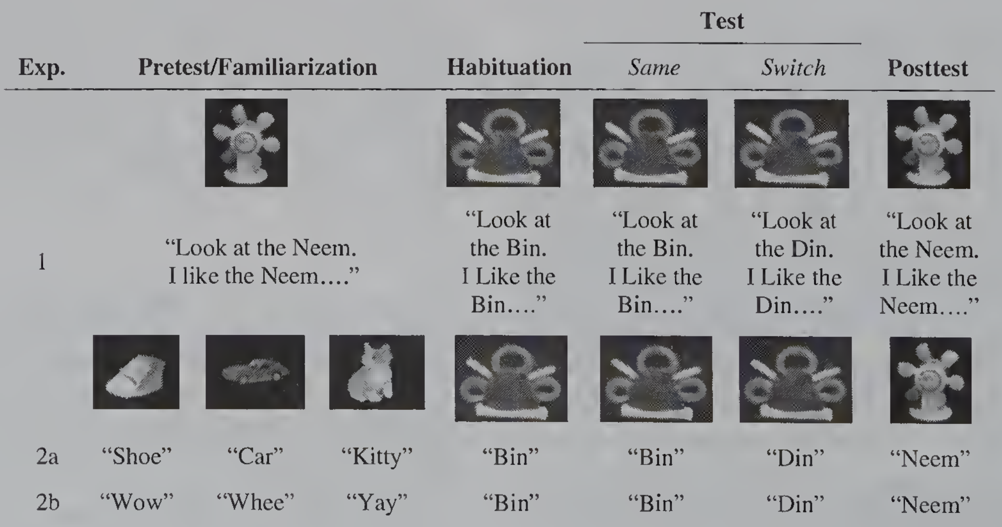

Test
Exp.
Pretest/Familiarization
Habituation
Same
Switch
Posttest
1
"Look at the Neem. I like the Neem…."
"Look at the Bin. I Like the Bin…."
"Look at the Bin. I Like the Bin…."
"Look at the Din. I Like the Din…."
"Look at the Neem. I Like the Neem…."
2a
"Shoe"
"Car"
"Kitty"
"Bin"
"Bin"
"Din"
"Neem"
2b
"Wow"
"Whee"
"Yay"
"Bin"
"Bin"
"Din"
"Neem"

Figure 1. Experimental design and stimuli.

using dissimilar stimuli from habituation (the waterwheel–*neem* pairing), followed the test phase. Infants recovered from the last habituation block to this posttest in both Experiments 1 and 2 (all *p* values < .01), ensuring that infants were not fatigued or generally disinterested in the task. Using a frame-by-frame analysis (1 frame = 33.33 ms), two experienced coders, blind to condition assignment, scored infants' looking times (*r* > .95 in Experiments 1 and 2).

Results and Discussion

When the novel words were embedded within naming phrases, 14-month-olds successfully harnessed their finely tuned phonetic sensitivities to map that word to its intended referent. See Figure 2. Infants looked reliably longer on switch (*M* = 11.67 s, *SD* = 4.56) than on same (*M* = 9.46 s, *SD* = 3.88) trials, *t*(15) = 2.36, *p* = .03, *d* = 0.52.

Infants' success in this task is striking. In contrast to infants hearing isolated words, these infants had to parse the novel words from the continuous speech stream. Their success, even in the face of this more crowded perceptual space, suggests that embedding the novel words in phrases did indeed clarify their referential status for the infants. But could infants' success be attributed to something else? In continuous speech, each phoneme (e.g., *b*) is influenced by those surrounding it; the transitions between phonemes provide coarticulatory cues, to which infants as young as 5 months show sensitivity (Fowler, Best, & McRoberts, 1990). Perhaps by embedding novel words in phrases, we provided infants in Experiment 1 with coarticulatory cues (transitions between determiners and novel words) that were unavailable to infants hearing isolated words in previous versions of the Switch task. To discover whether infants' success in Experiment 1 could be attributed to this additional perceptual information or to the referential status of the novel words, we designed a second experiment.

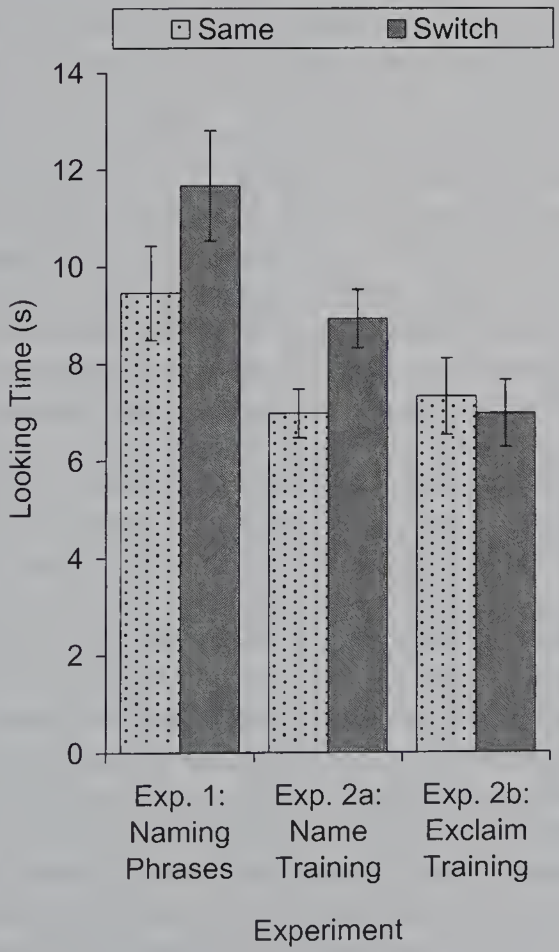


Figure 2. Mean looking times (and standard error) for same and switch trials across conditions.

Experiment 2

Our goal was to present the novel words in isolation, as in previous versions of the Switch task, but to maintain their referential clarity. To achieve this, we took advantage of a compelling finding: If infants first have an opportunity to observe a series of familiar objects (e.g., car, kitty), each paired with its familiar basic level name presented in isolation (e.g., "Car!" "Kitty!"), they subsequently map *novel* words, also presented in isolation, to novel objects. But if the very same familiar objects are initially paired with familiar exclamations (e.g., "Wow!" "Whee!"), infants fail to establish consistent mappings (Namy & Waxman, 2000).

Building upon this insight, we introduced a brief training period in which infants were introduced to three familiar objects (car, shoe, cat). Infants were randomly assigned to one of two training conditions. In the name-training condition, each training object (e.g., car) was paired with its familiar basic-level object name, presented in isolation (e.g., "Car!"). In the exclaim-training condition, each training object was paired with a familiar exclamation (e.g., "Wow!"). We then observed infants' performance in the Switch task, with novel words presented in isolation.

If infants' success in Experiment 1 can be attributed to perceptual factors (e.g., coarticulatory cues), then infants in both training conditions should fail to map the novel words, now presented in isolation, to the accompanying object in the subsequent Switch task. But if instead, infants' success in Experiment 1 is a consequence of increased referential clarity, then performance in the name- and exclaim-training conditions should differ. We predict that the name-training manipulation will establish that isolated words can refer to objects in the context of this task and that, as a result, infants in this condition will map the novel word in all its detail to the accompanying object. In contrast, although infants in the exclaim-training condition also hear familiar words, and although these words are themselves attention enhancing (e.g., ''Wow''), they do not refer specifically to objects. Therefore, we predict that infants in the exclaim-training will not infer that isolated words function as names for objects in this task and, as a result, will not map the novel word to the accompanying object in the subsequent Switch task.

Method

Participants

Twenty-eight English-learning, healthy, full-term infants (*M* = 14.37 months, range = 13.55–15.49 months) from greater Chicago participated. All infants comprehended the basic-level name for at least one training object, as measured by the MCDI. An additional 16 infants were excluded due to restlessness (8), comprehending none of the training object labels (5), technical error (1), or parental interference (2). There were no differences between the two conditions on any vocabulary measure (see Table 1).

Stimuli

The novel words, identical to those in Experiment 1, were now presented in isolation. The training phase included moving images of a car, cat and shoe with their basic-level names (*car*, *kitty*, *shoe*) or with familiar exclamations (*whee*, *wow*, *yay*).

Procedure

This was nearly identical to Experiment 1 with two exceptions (see Figure 1). First, in lieu of the pretest, we introduced a training phase: Infants saw three familiar objects, presented sequentially, each paired with either its appropriate basic-level name or an exclamation. The order in which these training trials were presented was counterbalanced across infants within training conditions. Second, because novel words were presented in isolation, habituation trials were reduced from 20 to 14 s in duration (similar to Stager & Werker, 1997).

Table 1
Experiment 2: Vocabulary Data

Vocabulary measures	Condition: Name	Condition: Exclaim	*t*(26)	*p*
Vocabulary: Comprehension	153.64	127.29	0.96	0.35
Vocabulary: Production	15.00	17.86	0.45	0.66
Training words comprehended	2.50	2.64	0.54	0.59
Training words produced	0.43	0.50	0.29	0.77

Results and Discussion

Even in the absence of any coarticulatory cues, infants in the name-training condition successfully established word–object mappings; those in the exclaim-training condition did not. A 2 (test trial: same vs. switch) by 2 (condition: name- vs. exclaim training) mixed analysis of variance revealed no main effects: trial, $F(1, 26) = 2.43$, $p = .13$, partial $\eta^2 = 0.08$, and condition, $F(1, 26) = 1.06$, $p = .31$, partial $\eta^2 = 0.04$. A significant Trial × Condition interaction, $F(1, 26) = 5.02$, $p = .03$, partial $\eta^2 = 0.16$, revealed the predicted outcome (Figure 2). Infants in the name-training condition established a word–object pair, looking reliably longer on switch ($M = 8.92$ s, $SD = 2.26$) than on same ($M = 6.97$ s, $SD = 1.87$) test trials, $t(15) = 2.53$, $p = .02$, $d = 0.94$. Thus, infants' use of fine phonetic detail in Experiment 1 cannot be attributed to low-level perceptual factors alone. In contrast, infants in the exclaim-training condition failed to detect the switch, $t(15) = 0.52$, $p = .61$, $d = 0.13$.

The difference between these two conditions cannot be attributed to selective adaptation, a phenomenon in which repeated exposure to one endpoint of a phonetic continuum (e.g., [b]) leads to a perceptual bias favouring the other endpoint (e.g., [d]). We tallied the number of habituation tokens (*bin* or *din*; 7 per trial) presented to infants in each of the conditions and found no significant difference (name: $M = 51$, $SD = 24.32$; exclaim: $M = 73$, $SD = 45.67$), $t(26) = -1.59$, $p = .12$, $d = -.60$. Furthermore, there was no correlation between this measure and infants' increased looking time to switch, $r(26) = 0.09$, $p = .66$. Moreover, an analysis of infants' vocabularies provided additional evidence that referential, and not perceptual, factors, accounted for infants' success. We tallied the number of training words comprehended by each infant (0–3), including those 5 infants who were excluded from the preceding analyses because they comprehended none of the training objet labels. This measure was correlated with success in the name-training condition, where comprehension was crucial to understanding that isolated words could refer to the objects, but not the exclaim-training condition: name, $r(16) = .59$, $p = .01$, and exclaim, $r(13) = -.24$, $p = .38$.

General Discussion

Together, these experiments constitute strong support for the hypothesis that 14-month-old infants can indeed recruit their finely honed sensitivities to phonetic detail in the task of word learning, and do so effectively when the referential status of the to-be-learned novel word is clear. When novel words are presented in an unambiguously referential context (naming phrases in Experiment 1; name training in Experiment 2), 14-month-olds recruit their perceptual sensitivities to establish precise word–object mappings. When the referential status of the novel word remains ambiguous (as in the original Switch tasks; Experiment 2: exclaim training), infants fail.

Moreover, these results take us one step further, demonstrating that infants have multiple ways to identify referential status. In Experiment 1, where the novel words were embedded in naming phrases, infants took advantage of syntactic cues to establish their referential status. In Experiment 2, where no syntactic information was provided, infants capitalized on pragmatic cues to establish referential status (Namy & Waxman, 2000). Infants in the name-training condition, who were provided with pragmatic evidence that words presented alone could indeed be names for objects, were able to exploit this evidence to map novel words to objects in the subsequent Switch task and to use fine phonetic detail in doing so. Importantly, infants' persistent failure to do so in the exclaim-training condition documents that simply presenting infants with familiar objects paired with familiar attention-directing words is not sufficient. Instead, what 14-month-old infants require to establish a word–object mapping is a reason to expect that the word is meant to refer. Although it is beyond the scope of the current article to consider fully the factors involved in infants' emerging understanding of reference (see Waxman & Gelman, 2009), it is clear that 14-month-old infants are sensitive to referential cues, that these support infants' establishment of word–object mappings, and permit infants to bring their finely honed phonetic sensitivities into play in the process of word learning. Thus, bringing in the notion of reference—a fundamental feature of human language—helps to resolve what was previously seen as a paradox in infant word learning.

Our proposal concerning the essential role of referential status also brings together a series of otherwise apparently unrelated recent findings under one umbrella rather than appealing to a diverse array of methodological and perceptual explanations. For example, when 14-month-olds are introduced to a novel object in the context of a focused, social interaction with an adult, they attend successfully to fine phonetic detail when establishing word–object mappings (Mani & Plunkett, 2008). We suspect that this animated social interaction, focusing specifically on an object for which infants had no known name, served to increase their interest in discovering its name. Similarly, when 14-month-olds are permitted to play with an unnamed novel object in their homes in the weeks preceding their lab visit, they attended to phonetic detail when mapping a novel word to that object in the lab during the Switch task (Fennell, 2004). Here, we suspect that infants' prolonged engagement with an unnamed toy augmented their interest in discovering its name when provided with candidates in the laboratory during the Switch task. Evidence from another recent report fits well with our focus on reference and meaning. When novel words were produced by a range of *different* speakers during habituation, 14-month-olds successfully detected changes in phonetic detail (Rost & McMurray, 2009). Rost and McMurray (2009) attributed infants' success to perceptual factors, hypothesizing that exposure to a range of exemplars supported infants' ability to build stable phonetic categories. We concur but point out that this task likely offered infants additional referential information as well. After all, when a range of different speakers consistently applies the very same word to a novel object, this social convergence signals that that word is the name of that object.

Another recent report indicates that clarifying referential status in the testing phase also aids infants. Yoshida et al. (2009) presented a version of the Switch task in which 14-month-old infants were introduced to two similar-sounding words (*bin* and *din*), each paired with a unique novel object. At test infants saw both novel objects simultaneously but heard only one of the two novel words. To succeed, infants had to map each of two similar-sounding words presented during habituation to its associated object and to represent these words with sufficient phonetic detail to direct their attention to the correct object at test. Surprisingly, 14-month-olds succeeded. Yoshida et al. argued that presenting the two novel objects simultaneously at test reduced the memory demands of the task, leading to infants' success. We take no issue with this interpretation but point to another design feature that was likely instrumental. To familiarize them with the forced-choice nature of this task, infants viewed ''filler trials'' before viewing the test trials themselves. The structure of the filler trials was nearly identical to that of the test trials, with one important exception: Filler trials included objects and words *familiar* to the infants (e.g., they saw a car and shoe side by side and heard *car*). In our view, these filler trials were instrumental not only because they oriented infants to the structure of the test trials but also because the inclusion of familiar words and objects, (like the name-training trails of Experiment 2) served to establish the referential status of novel words.

Each of these recent demonstrations, considered on its own, sheds light on a distinct factor mediating infants' performance. But invoking the notion of referential status provides a more integrative account for infants' patterns of successes and failures. In our view, these recent findings converge to support the proposal that infants use social, linguistic and pragmatic information as windows into the referential status of a novel word, and that once this status is assured, they recruit their sensitivity to fine phonetic detail to map that word to its intended referent.

Although we have focused expressly on the essential role of referential status in word learning, we acknowledge the importance of other elements as well, chief among them the perceptual factors. Recent evidence establishing the power of perceptual cues comes from Thiessen (2007). Before participating in a Switch task, infants in this experiment were exposed to a training period in which the novel words were presented under variable phonetic contexts. This purely perceptual boost in training supported infants' subsequent use of phonetic detail in the Switch task, suggesting that when the phonetic distinction is sufficiently robust (as with the completely distinct words *lif* and *neem*), 14-month-olds succeed (Werker et al., 1998). Infants' success in Thiessen (2007) and Werker et al. (1998) reveals that even in the absence of referential clarity, infants' attention to phonetic detail can be augmented by perceptual means. Importantly, the results of the current experiments reveal that the corollary is also true: Even in the absence of any added perceptual information, infants' attention to phonetic detail is boosted when the referential status of the word is made clear. In future work, it will be important to discover how various factors (perceptual, syntactic, social, and pragmatic) come together in the service of word learning.

The current work provides a resolution to what has been seen as a paradox. It reveals that there is considerable continuity in the phonetic representations that infants refine over the 1st year of life and the lexical representations they establish thereafter. It also fortifies the view that infants (like adults) take into consideration more than associations alone when establishing word meaning (Gogate & Bahrick, 1998; Waxman & Lidz, 2006). In every version of the Switch task, the presence of a novel word is correlated perfectly with the presence of a novel object. Yet this correlation alone, however perfect, is not sufficient to support the establishment of a strong enough word–object mapping to reveal infants' use of fine phonetic detail. We therefore conclude that when mapping words to meaning, infants recruit both their sensitivity to fine phonetic detail and their sensitivity to the distinct referential status of words.

References

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 117–134). Mahwah, NJ: Erlbaum.

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, *80*, 15–22.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, B33–B44.

Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 67–112). Amsterdam: John Benjamins.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Fennell, C. T. (2004). *Infant attention to phonetic detail in word forms: Knowledge and familiarity effects* (Unpublished doctoral dissertation), University of British Columbia, Vancouver, Canada.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, 1–173.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*, F33–F40.

Fisher, C., Church, B., & Chambers, K. E. (2004). Learning to identify spoken words. In D. G. Hall & S. R. Waxman (Eds.) *Weaving a lexicon* (pp. 3–40). Cambridge, MA: MIT Press.

Fowler, C. F., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, *48*, 559–570.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, *105*, 218–228.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*, 133–149.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language perception between 6 and 12 months. *Developmental Science*, *9*, F1–F9.

Mani, N., & Plunkett, K. (2008). 14-month-olds pay attention to vowels in novel words. *Developmental Science*, *11*, 53–59.

Namy, L. L., & Waxman, S. R. (2000). Naming and exclaiming: Infants' sensitivity to naming contexts. *Journal of Cognition and Development*, *1*, 405–428.

Pater, J., Stager, C. L., & Werker, J. F. (2004). The lexical acquisition of phonological contrasts. *Language*, *80*, 361–379.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*, 432–443.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*, 16–34.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*, B35–B50.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*, 217–242.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*, 258–263.

Waxman, S. R., & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology*, (6th ed., Vol. 2, pp. 299–335). Hoboken, NJ: Wiley.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month old infants. *Developmental Psychology*, *34*, 1289–1309.

Werker, J. F., & Fennell, C. T. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. Waxman (Eds.), *Weaving a lexicon*. (pp. 79–109). Cambridge, MA: MIT Press.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary. *Infancy*, *3*, 1–30.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Yoshida, K., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar sounding words. *Developmental Science*, *12*, 412–418.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1384–1399

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective

Maike Gieling, Jochem Thijs, and Maykel Verkuyten
Utrecht University

Using social-cognitive domain theory and social identity theory, tolerance judgments of practices by Muslim actors among Dutch adolescents (12–17) were investigated. The findings for Study 1 ($N = 180$) demonstrated that participants evaluated 4 practices using different types of reasons: personal, social-conventional, and moral. In Study 2 ($N = 970$), participants were more tolerant of acts considered to be a personal issue than of acts that pertained to moral issues (social-conventional acts were in the middle). Furthermore, participants were more tolerant of the particular practices than of campaigns for public support of co-Muslims for these practices. Level of education, in-group identification, and multiculturalism had much weaker effects in the moral than in the other domains. Age was negatively related to tolerance.

Islam has moved to the center of debates and politics in European countries and is at the heart of what is perceived as a ''crisis of multiculturalism'' (Modood & Ahmad, 2007; Zolberg & Long, 1999). The hotly debated issues are about concrete actions and behaviors. Should the practice of forced marriages among some Muslim groups be accepted? Should it be accepted that Muslim teachers refuse to shake hands with children's parents of the opposite sex? Should it be allowed that civil servants wear a headscarf and that students wear a burqa or a niqab? It is around these concrete questions that multiculturalism is put to the test and ways of life can collide. Studies on political thinking and behavior show that individuals tend to support democratic rights in the abstract but often do not endorse the same rights in concrete circumstances (see Vogt, 1997). It is one thing to endorse the freedom of speech and another thing to apply this freedom to, for example, an imam calling homosexuals inferior people.

The present article reports on two studies examining ethnic majority Dutch adolescents' tolerance of negatively evaluated practices by Muslim actors. Theoretically, the research aims to bring together social-cognitive domain theory (see Killen, Margie, & Sinno, 2006; Smetana, 1995; Turiel, 2002) and intergroup theories (see Nesdale, 2008; Tajfel & Turner, 1979). We examined tolerance judgments related to practices that evoke personal, social-conventional, and moral considerations and we studied the role of in-group identification and the endorsement of multiculturalism. Thus, we employed an integrative social-developmental perspective that considers social-psychological processes together with developments in cognitive knowledge and reasoning (Killen & McKown, 2005; Levy, West, & Ramirez, 2005; Pomerantz & Newman, 2000).

We conducted our research with ethnic majority Dutch participants between 12 and 17 years of age and we also examined the effects of age, gender, and level of education. In the Netherlands, the level of prejudice toward Muslims is relatively high among the general population and also among adolescents (Pew Global Project; Pew Research Center, 2005; Velasco Gonzalez, Verkuyten, Weesie, & Poppe, 2008). According to some commentators, there is an ongoing ''Dutch–Muslim'' cultural war and a related culture of fear (Scroggins, 2005; Vasta, 2007). Leading politicians have taken a fiercely negative position on Islam, which is defined as a backward religion that seriously threatens Dutch society, and national identity and culture (Verkuyten & Zaremba, 2005). Thus, our research examines tolerance in a rather negative intergroup setting.

Diversity and Diversity

There are different kinds of diversity and some are more problematic than others. For instance,

---

Correspondence concerning this article should be addressed to Maike Gieling, Faculty of Social Sciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, Netherlands. Electronic mail may be sent to m.gieling@uu.nl.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0006

when people differ in their moral beliefs it is difficult to build a cohesive community. Take the example of the acceptance of homosexuality. The acceptance of homosexuality is a matter of personal opinion when one accepts it oneself but does not expect others to do so. It becomes conventionally regulated when one argues that everyone in one's own group or society should accept it but people in other groups or societies need not. But if one sees the acceptance of homosexuality as a moral issue one thinks that everyone else in all communities should support it, for example, as a basic human right. In that case, it becomes difficult to tolerate that others think differently.

This distinction is similar to the one proposed in social-cognitive domain theory (Killen, Margie, & Sinno, 2006; Smetana, 1995; Turiel, 2002). The social-cognitive domain model emphasized that children and adolescents apply different domains of knowledge in their social reasoning and judgments. Not only moral considerations (fairness, welfare, and rights) but also social-conventional (group norms and traditions) and personal ones (individual autonomy) are applied to a range of social events. Studies on children's social reasoning indicate that these domains emerge independently and simultaneously in early development (Kim & Turiel, 1996; Laupa & Turiel, 1993; Nucci, 2001). Children distinguish between them and react differently to personal considerations, conventional norms, and moral norms.

The distinction between these domains has been criticized because it would be cultural specific (Shweder, Mahapatra, & Miller, 1987). In addition, it has been argued that people often treat moral harms and injustices as merely conventional violations and sometimes react with moral outrage to what look like harmless violations of conventions (Kelly, Stich, Haley, Eng, & Fessler, 2007). However the domain theory claims that it is the personal, conventional, and moral distinction that appears at a young age and research has provided extensive cross-culture validation (see Turiel, 1998; Wainryb, 2006). Furthermore, because moral actions have intrinsic consequences for others' rights and welfare it is claimed that there is substantial cross-cultural agreement about what is moral, whereas there is more cultural variation in the content of conventional and personal issues. One implication is that the meaning of particular practices or behaviors is not always self-evident. The different interpretations of homosexuality mentioned earlier, is an example. Thus, to understand adolescents' differential judgments about various Muslim practices it should first be examined whether they make a distinction between these practices in terms of the personal, the conventional, and the moral domains.

Study 1 sets out to do this. We used four concrete cases of specific practices that are not illegal but that are hotly debated in Dutch society: the wearing of a headscarf by Muslim women, the refusal to shake hands with males by a female Muslim teacher, the founding of separate Islamic schools, and the public expression of the view that homosexuals are inferior people by an imam. These cases can reflect considerations that pertain to different domains at the same time and in that respect can be thought of as mixed-domain or multifaceted issues (Turiel, 1983, 1998). However, in comparison to each other, they are expected to differ in the relative salience of a particular type of domain that they evoke. For example, calling homosexuals inferior people can cause psychological harm to others because it is viewed as an attack on their way of being. This reflects moral considerations and could therefore be considered primarily a "moral" issue. No research has examined how individuals examine the wearing of a headscarf by a Muslim. The latter may be more likely to be considered a personal choice of the female herself given the apparent absence of harm to others or denial of resources (however, being denied the right to wear a headscarf may be viewed as "moral"). Not shaking hands with people from the opposite sex and the founding of Islamic schools were expected to be evaluated using predominantly social-conventional reasoning, as both deal with societal expectations. Thus, in Study 1 we predicted that adolescents will make a distinction between the four cases in terms of concerns predominantly raised by the three domains.

Subsequently, in Study 2 we examined whether tolerance judgments differ for the three domains. Specifically, following domain theory we expected adolescents to be less tolerant of practices that are perceived to violate a moral norm compared to social-conventional violations. In addition, practices that are evaluated negatively but considered as "personal" should be tolerated the most.

Practices and Social Support

Accepting that people hold dissenting beliefs does not have to imply that one tolerates the public expression of such beliefs or the actual practices based on such beliefs (Vogt, 1997). These dimensions of tolerance can trigger different levels of endorsement. In their study, Wainryb, Shaw, and Maianu (1998) found, for example, that European American children and early adolescents were more tolerant of dissenting speech than practices (see also Witenberg, 2002). Verkuyten and Slooter (2007) found that Dutch adolescents were more tolerant of Muslim parents publicly arguing for differential gender treatment or for a very light form of female circumcision than for the actual acts themselves. This higher acceptance of the public expression of the dissenting beliefs is consistent with the idea of free speech. It can be seen as stimulating debate that is important for the democratic process and as causing less direct harm or injustice than the actual acts. This means that it could be expected that adolescents are relatively more tolerant of public expressions of Muslim beliefs than of the actual practices based upon these beliefs.

However, this might depend on the intergroup context and the aim of the public expression. Specifically, Muslims trying to persuade cobelievers to engage in a dissenting practice can be perceived as a threat to the Dutch in-group. Ethnic majority Dutch adults and adolescents tend to see Muslims as threatening the Dutch way of life and this has been found to lead to strong negative and intolerant reactions (Van der Noll, Verkuyten, & Poppe, 2010; Velasco Gonzalez et al., 2008). Further, the Dutch majority tends to reject particular practices of Muslims and a recent nationwide survey showed that 50% of the Dutch consider the Western and Muslim way of life as opposites that do not go together (Gijsberts, 2005). In Study 2, the focus is not only on the participants' tolerance of the different practices but also on their acceptance of people trying to mobilize fellow Muslims. Participants were asked whether it should be allowed that the different Muslim actors campaign in order to try to convince others to do the same thing. This social mobilization of Muslims is typically seen as threatening to Dutch identity and culture (Sniderman & Hagendoorn, 2007; Velasco Gonzalez et al., 2008), and therefore the adolescents were expected to be less tolerant of Muslims campaigning for in-group support for the particular practice than of the actual practice itself.

In-Group Identification

Social identity theory (SIT; Tajfel & Turner, 1979) argues that individuals seek to belong to groups that provide them with a secure and positive social identity and are motivated to maintain positive distinctiveness through intergroup comparisons. Establishing favorable evaluative distinctiveness of one's group vis-à-vis other groups helps to achieve a positive group identity. However, negative out-group evaluation is by no means an automatic product of group distinctions, but a function, for example, of the extent to which children identify with their social group, the norms and beliefs held by the members of the child's social group, and the extent to which the out-group is perceived as threatening one's group in some way (Nesdale, 2008).

SIT interprets group identification in terms of individual differences in the degree to which psychologically central and valued group memberships develop. Some children are more inclined than others to see themselves as an ethnic group member and to value their ethnic group membership. Children who feel highly committed to their group are inclined to act in terms of their group membership. The theory argues that group identification interacts with out-group threat to predict out-group evaluation and tolerance of minority practices. The idea is that perceived threat has different effects depending on in-group identification because the motivational meanings of perceived threat are different. Compared to low identifiers, those with high in-group identification are more likely to be concerned about their group, especially when the position and value of the group identity are at stake. For example, high identifiers have been found to react more negatively toward the out-group under threat than do low identifiers (e.g., Bizman & Yinon, 2001; Tausch, Tam, Hewstone, Kentworthy, & Cairns, 2007). Thus, we expected that Dutch adolescents who feel strongly committed to their group will be less tolerant of Muslim practices compared to adolescents with relatively low in-group identification. In addition, we examined whether group identification equally affects tolerance judgments in the personal, the conventional, and the moral domain. Abrams, Rutland, Ferrell, and Pelletier (2008) found that children who identified more with their group showed in-group favoritism for conventional-based judgments but less so for morality-based judgments. Moral criteria are general and apply to in-group and to out-group behavior. Therefore, they should not be affected strongly by in-group identification. Thus, in Study 2, we expected higher in-group identifiers to be less tolerant of the practices and of social mobilization than lower identifiers, for acts that are considered more personal and social-conventional, but not for acts that raise moral concerns.

Multiculturalism

Individuals differ in the extent to which they endorse and value diversity within a society. A multicultural ideology entails "the general and fundamental view that cultural diversity is good for a society and for its individual members and that diversity should be shared and accommodated in an equitable way" (Berry, 2006, p. 728). Experimental research has shown that participants exposed to a message that endorses multicultural recognition show more positive out-group attitudes than those exposed to a message that supports a color-blind or assimilationist perspective (e.g., Verkuyten, 2005; Wolsko, Park, Judd, & Wittenbrink, 2000). Furthermore, and in relation to Muslims in the Netherlands, multiculturalism seems to provide a general ideological view about the importance of cultural diversity that reduces a sense of group threat. It also emphasizes that minorities should be recognized and valued in their group identity, and that there should be equal opportunities and rights (Van der Noll et al., 2010; Velasco Gonzalez et al., 2008). Therefore, we expected in Study 2 a positive association between the endorsement of multiculturalism and tolerance. In addition, we expected this association for the practices in the personal and the social-conventional domain but, again, not in the moral domain. The endorsement of cultural diversity is much more difficult when it concerns issues of harm, justice, and fairness.

Age, Education, and Gender

Although several studies have described a developmental progression from a generally intolerant attitude during the childhood years to increasingly tolerant judgments during adolescence (e.g., Bobo & Licardi, 1989; Enright & Lapsley, 1981; Enright, Lapsley, Franklin, & Steuck, 1984), different aspects of tolerance and domains of dissenting beliefs and practices are not taken into account in this research. Studies that do examine these different aspects present a more complex picture of age differences with tolerance and intolerance coexisting at all ages (e.g., Verkuyten & Slooter, 2007; Wainryb, Shaw, Laupa, & Smith, 2001; Wainryb et al., 1998). In contrast to young children, (early) adolescents tend to make more context-sensitive judgments about freedoms and rights. Thus, with age, concepts of rights and freedoms are increasingly evaluated in relation to other considerations and concerns, including the intergroup context (e.g., Helwig, 1995, 1997; Ruck, Abramovitch, & Keating, 1998). For Study 2, this might mean that older adolescents are less tolerant because they are more aware of the public debate about the Islamization of Dutch society. Lesser tolerance among older adolescents should exist for the social-conventional domain in particular, but not as clearly for the moral domain.

Chronological age is often used as an indicator for the level of cognitive and moral development. Age differences are typically interpreted in terms of increased cognitive skills and knowledge. The importance of these factors in explaining tolerance judgments is also suggested by the positive association between education and different forms of tolerance. In his review, Vogt (1997) concludes that there is clear evidence that education increases tolerance. Both the number of years of education and the level of education seem to contribute to this. In Study 2, and in addition to age, we examined the level of education. It was expected that a higher level of education has a positive effect on tolerance, but, again, less so for tolerance in the moral domain.

Several studies have found gender differences in tolerant judgments. For example, Killen, Sinno, and Margie (2007) showed that boys compared to girls more often found ethnic and racial exclusion acceptable. Verkuyten and Slooter (2007) also reported gender differences, with male adolescent being less tolerant than females. However, other studies using cognitive domain theory have found few gender differences (Smetana, 2006) or no differences at all (e.g., Helwig, 1995, 1997; Sigelman & Toebben, 1992). In Study 2, we explored possible gender differences in tolerance of Muslim practices.

Summary

The main purpose of this research was to examine tolerance of practices by Muslim actors among Dutch adolescents by applying social-cognitive domain theory to an intergroup context as has been done in prior research (Abrams et al., 2008; Killen et al., 2007). The novel approach taken in this study is to examine social reasoning together with group identity and the endorsement of multiculturalism in the Dutch–Muslim context, and to focus on aspects of tolerance regarding societally controversial Muslim practices. We focused on concrete cases that are debated in Dutch society and reflect different aspects of tolerance. In Study 1, we first examined empirically whether these cases can be distinguished in referring predominantly to personal, social-conventional, and moral concerns and considerations. In Study 2, we examined tolerance judgments in relation to these three domains and to in-group identification and the endorsement of multiculturalism.

Study 1

The purpose of the first study was to examine the type of concerns and considerations that are evoked by the different Muslim practices. Based on national public debates in the media, we first developed eight short descriptions of controversial cases involving Muslim actors. In a small pilot study, these eight cases were discussed with 10 adolescents. Four of the eight cases were found to be evaluated rather negatively and thus raised questions of tolerance (Vogt, 1997). These four cases reflected aspects of each social domain: the personal, the social-conventional, and the moral. Thus, we used these four cases and examined among a larger sample whether these cases did indeed indicate the three domains.

Method

Participants

The sample included 180 ethnically Dutch participants (''White'' majority and non-Muslim) between 12 and 17 years of age ($M = 14.12$, $SD = 1.73$). In total, 43% were males and 57% were females. Of the participants, 11.7% were in the lowest level of secondary education (junior vocational track), 18.3% in the middle level (higher vocational track), and 33.3% in the highest level of education (pre-university track). An additional 30% of the participants were in a mixed middle/higher level track. The pupils participated on a voluntary basis and the anonymous paper-and-pencil questionnaire was administered during regular class hours and under supervision.

Measures

The participants were asked to judge four different cases. The four cases involved: (a) the wearing of a headscarf (''Imagine that a group of students in your school wears a headscarf which covers a small part of the face''), (b) not shaking hands with someone of the opposite sex (''Recently, there has been a discussion about a female Muslim teacher who does not want to shake hands with men, such as colleagues and fathers of students''), (c) the founding of separate Muslim schools (''Some parents are founding Islamic schools where only Muslim children are accepted''), and (d) the voicing of harmful opinions about homosexuals (''Recently, an imam held a speech in a mosque in which he calls homosexuals inferior people''). The first case was expected to be predominantly considered a personal matter, the second and third were expected to mainly raise social-conventional concerns, and the fourth was expected to be considered a violation of a moral norm.

Modifications of standard social domain assessments were applied in this study to measure participants' interpretation of each case (see Smetana, 2006, for standard assessments). We investigated the type of judgments along three dimensions, each reflecting a different aspect of the conventional, the personal, and the moral domains, respectively. The first dimension entailed the extent to which the act is regarded as unconventional, and the second assessed whether the act is under personal jurisdiction of the actor (a criterion for the personal domain). The third dimension involved intrinsic harmful consequences for others, which is a justification category for using the moral domain (Smetana, 2006; Turiel, 1983). In addition, we included a question on the perceived wrongness of the act. Wrongness and domain type are conceptually independent, but moral transgressions tend to be perceived as more wrong than social-conventional violations (Smetana, 2006; Tisak & Turiel, 1988). Hence, wrongness ratings may be used to obtain additional support for the notion that an act elicits moral concerns.

For each case, the questions had the same response format. With respect to the issue of the headscarf, for example, students had to answer the following four questions: ''Is it inappropriate to wear headscarves in Dutch schools?'' (unconventional), ''Is it a personal decision to wear a headscarf?'' (personal), ''Are other people in school in any way harmed by it?'' (harmful consequences), and ''Is it wrong to wear a headscarf?'' (wrongness). The harmful consequences measure had a 5-point scale with answers ranging from −2 (*not at all*) to 2 (*very much*). The other measures were 7-point scales ranging from −3 (*no*) to 3 (*yes*).

Results

Both relative and absolute criteria were used to examine whether students assessed the four cases differently. First, for each separate judgment and for each case, we calculated the percentages of agreeing and disagreeing participants, and we conducted *t* tests to examine whether mean scores exceeded the neutral scale midpoint. Next, to test whether the scores on each of the dimensions differed between the practices, we conducted four separate repeated measures analyses of variance (ANOVAs), with the type of practice as the repeated measures factor. Results are shown in Table 1.

The absolute criteria revealed that all cases were generally considered as nonconventional. The mean scores on the nonconventional dimension were significantly higher than zero, and the majority of participants selected answers on the positive side of the scale. Yet, the ANOVA revealed a significant main effect of type of practice which indicated that some cases were considered more nonconventional than others, $F(3, 483) = 50.769$, $p < .001$, $\eta_p^2 = .24$. Pairwise comparison (LSD) showed that this was most strongly the case for the refusal to shake hands with someone of the opposite sex, followed by the imam's speech. The founding of Islamic schools and the wearing of a headscarf scored the lowest on the nonconventional dimension.

Mean scores and (dis)agreement percentages for the personal dimension revealed that the wearing of a headscarf was clearly considered a matter of personal choice, whereas the founding of Islamic schools and the voicing of harmful opinions about homosexuals were clearly not. Students were "undecided" about the refusal to shake hands. ANOVA results were consistent with this pattern, $F(3, 486) = 23.79$, $p < .001$, $\eta_p^2 = .13$.

Table 1
Mean Scores and Percentage Agreement for Judgment Dimensions by Case: Study 1

Dimension	Headscarf	School	Hand	Speech
Nonconventional				
M (*SD*)	0.62 $(1.88)^{**}_{a}$	0.93 $(1.68)^{**}_{a}$	2.33 $(1.13)^{**}_{c}$	2.01 $(1.56)^{**}_{b}$
% agree	53.5	59.2	91.0	81.0
% disagree	22.9	20.7	1.8	6.7
Personal				
M (*SD*)	1.01 $(2.02)^{**}_{a}$	−0.51 $(2.01)^{**}_{c}$	0.15 $(2.11)_{b}$	−0.46 $(2.25)^{*}_{c}$
% agree	62.8	29.3	41.2	31.3
% disagree	26.2	55.7	42.9	54.6
Harmful consequences				
M (*SD*)	−0.34 $(1.26)^{**}_{a}$	−0.45 $(1.06)^{**}_{a}$	0.22 $(1.27)^{*}_{b}$	1.43 $(0.82)^{**}_{c}$
% agree	31.6	16.3	47.2	88.2
% disagree	45.0	43.0	24.5	1.9
Wrongness				
M (*SD*)	−1.27 $(1.79)^{**}_{a}$	−0.14 $(2.05)_{b}$	−0.11 $(2.17)_{b}$	1.83 $(1.68)^{**}_{c}$
% agree	14.6	37.4	46.1	79.1
% disagree	62.6	42.0	38.9	9.8

Note. Means in the same row with different subscripts differ significantly at $p < .01$ by the Fisher least significant difference test.
$^{*}p < .05$. $^{**}p < .01$.

For the dimension of "harmful consequences," the mean scores suggest a distinction between the cases of the headscarf and school founding, on the one hand, and the cases of the handshaking and the imam's speech, on the other. The former cases were generally perceived as rather inconsequential, but the latter were considered to have a clear negative impact on others. However, this result is qualified by the (dis)agreement percentages and the ANOVA results, $F(3, 477) = 116.47$, $p < .001$, $\eta_p^2 = .42$. Although the score for the shaking of hands is above the scale midpoint, a minority of the respondents agreed that this practice has harmful consequences for others. Moreover, the mean score was significantly lower than that for the imam's speech ($p < .01$).

The results for "wrongness" indicate that participants clearly considered the speech against homosexuals as wrong, and the wearing of a headscarf as not wrong. Participants had a relatively neutral position in relation to the founding of Islamic schools and for the not shaking of hands. This pattern was confirmed by the ANOVA result, $F(3, 486) = 93.10$, $p < .001$, $\eta_p^2 = .37$.

To examine whether the scores for the dimensions were related to or qualified by age, education, and gender, we performed the same four ANOVAs with age included as a covariate and education and gender as factors. No significant main or interaction effects for these variables were found.

Discussion

The results of Study 1 revealed that participants differentiated between the four cases in terms of their ratings and responses, which reflected aspects of the personal, social-conventional, and moral domains. It is important to note that all practices were considered as nonconventional in the Netherlands. However, this does not mean that they exclusively evoked social-conventional concerns. As expected, the wearing of a headscarf was predominantly seen as an individual decision and a matter of personal choice. It was clearly not regarded as wrong or harmful for others. Both the not shaking of hands and the founding of Islamic schools were expected to be predominantly considered as social-conventional issues. The findings support this proposition. The majority of participants indicated that both practices were inappropriate in the Netherlands, whereas they did *not* judge them as being up to the actors, as having harmful consequences, or as being plainly wrong. Finally, the findings strongly support the idea that the imam's speech was predominantly considered a violation of moral norms. The majority of participants considered it wrong and harmful to make derogatory comments about homosexuals, and they clearly thought of it as a nonpersonal matter. Together, the findings provide support for the notion that these four cases tend to differ in the salience of the type of domains of sociomoral reasoning they evoke. To better understand participants' conceptualizations of these events, further analyses are required. In Study 2, we used these cases to examine tolerant judgments for the three domains and to test our hypotheses.

Study 2

The second study aimed to test four different hypotheses. First, participants were expected to be the least tolerant of the practice that for adolescents most strongly violates a moral norm (speech about homosexuals), followed by practices that predominantly violate conventional norms (the not shaking of hands and the founding of Islamic schools), and the most tolerant of the practice that relates most clearly to personal autonomy (the wearing of the headscarf). Next, it was expected that participants will be less tolerant of campaigns for the social support of fellow Muslims than of the actual practices themselves. Third, higher in-group identification was expected to be associated with less tolerance, whereas stronger endorsement of multiculturalism was expected to be associated with more tolerance. These two relationships were predicted for the practices that evoke personal and social-conventional considerations but not for the practice that is considered a moral transgression. Finally, we expected older adolescents to be less tolerant than younger ones, and we expected higher educated participants to be more tolerant. Again, these associations for age and education were expected to be weaker for the moral domain. In addition to testing these hypotheses, we explored the role of gender.

Method

Participants

This sample included 970 participants between 13 and 17 years of age ($M = 15.10$, $SD = 0.95$). In total, 51% were males and 49% were females. Of the participants, 36.3% were in the lowest level of secondary education (junior vocational track), 34.5% in the middle level (higher vocational track), and 29.2% in the highest level of education (pre-university track). At all three levels of education, all age groups participated. Thus, age and education were not confounded. Table 2 contains a breakdown of the sample by age, educational level, and gender. Both parents of the participants were ethnically Dutch ("White" majority and non-Muslim). The pupils participated on a voluntary basis and the anonymous paper-and-pencil questionnaire was administered in separate class sessions and under supervision.

Measures

The overall purpose of this study was to examine the contextual nature of adolescents' judgments about practices by Muslims. Therefore, two types of tolerance questions were used and because these two types of tolerance judgments have been found to be relatively independent (Verkuyten & Slooter, 2007), the measures were not counterbalanced but given in a fixed order. In order to measure the tolerance toward Muslim practices, the participants were asked to judge the four cases that were presented as follows: "Some parents are founding Islamic schools where only Muslim children are accepted. What should the government do about such schools?" "Imagine that a group of students in your school wears a headscarf which covers a small part of the face. What do you think the school should do?" "Recently, there has been a discussion about a female Muslim teacher who does not want to shake hands with men, such as colleagues and fathers of students. What do you think the school should do?" and "Recently, an imam held a speech

Table 2
Number of Respondents by Age, Educational Level, and Gender

Age group (years)	Educational level	Boys (%)	Girls (%)
13–14	Low	75 (7.7)	52 (5.4)
	Middle	21 (2.2)	24 (2.5)
	High	28 (2.9)	30 (3.1)
15	Low	66 (6.8)	54 (5.6)
	Middle	58 (6.0)	87 (9.0)
	High	85 (8.8)	88 (9.1)
16–17	Low	49 (5.1)	56 (5.8)
	Middle	86 (8.9)	59 (6.1)
	High	27 (2.8)	25 (2.6)

in a mosque in which he calls homosexuals inferior people. What do you think the mosque committee should do?''

In each case, there were four response categories: (a) ''Do nothing and allow it''; (b) ''Try to convince them [not to found those schools, not to wear a headscarf, to shake hands, to convince the imam to apologize], but allow it if they do not agree''; (c) ''Try to convince them not to do it, but forbid it if they do not agree''; and (d)''Simply not allow it.'' Items were recoded such that a higher score reflects a more tolerant attitude. In addition, to measure social support, the participants were asked if it should be allowed that the actors (i.e., the parents, students, teacher, imam) persuade other Muslims to engage in the same practice. Answers categories ranged from *definitely not allow it* (1), *to definitely allow it* (5). For each case, participants were also asked to rate their feelings toward the actors on a 7-point scale, ranging from *negative* (–3) to *positive* (3).

To measure *in-group identification*, participants were asked for their agreement with six different statements that have been used in various studies in the Netherlands (see Verkuyten, 2005). Three sample items are, ''My ethnic identity is an important part of myself,'' ''I strongly identify with my ethnic group,'' and ''I am proud of my ethnic background.'' Answers could range from *totally disagree* (1) to *totally agree* (5). Cronbach's alpha for the six items is .86. A higher score indicates stronger in-group identification.

The *endorsement of multiculturalism* was measured with eight items taken from Berry and Kalin's (1995) Multicultural Ideology Scale. These items have been used in previous research in the Netherlands (Arends-Tóth & van de Vijver, 2003; Verkuyten, 2006). Three sample items are: ''The more cultures there are, the better it is for the Netherlands,'' ''Immigrants in the Netherlands should forget their cultural background as soon as possible'' (reverse scored), and ''Migrants should be supported in their attempts to preserve their own cultural heritage in the Netherlands.'' Answers were given on 5-point rating scales: *strongly disagree* (1) to *strongly agree* (5). Cronbach's alpha is .81 and a higher score indicates a stronger endorsement of multiculturalism.

Analyses

The results are presented in three sections. The first section focuses on the tolerance judgments for the four cases, and the role of age, educational level, and gender. The second part describes differences for the two types of tolerance judgments, that is, tolerance of the practice itself and tolerance of the attempt to persuade others. The last section focuses on the effects of in-group identification and multiculturalism.

The concept of tolerance as we use it here implies forbearance or the putting up with something that one disapproves of or is prejudiced against (Vogt, 1997). Thus, in a preliminary analysis we first examined whether participants in general felt negative about the four practices. This turned out to be the case. Across the four cases, the average score lies well below the neutral midpoint (0) of the 7-point scales. One-sample t tests indicated that for each practice, the scores were significantly below zero, $ps < .001$, and even below −1, $ps < .001$. Thus, there were very few participants who indicated to have positive feelings and in general most participants in this sample were negative about the practices of the Muslim actors.

Results

Practices, Age, Education, and Gender

Tolerance judgments were expected to be dependent on the particular domain. We hypothesized that participants will be most tolerant of behavior that concerned mostly the personal choices of actors, and the least tolerant of moral violations. To examine these hypotheses, we first performed repeated measures ANOVA with type of practice as repeated measure. Next, we examined whether the pattern of tolerance judgments depended on students' educational level, gender, and age, by conducting 4 (type of practice) × 2 (gender) × 3 (educational level) repeated measures ANOVAs including age as a covariate, and type of practice again as the repeated measure. These analyses were conducted separately for tolerance of the actual practices and for tolerance of persuading fellow Muslims.

Tolerance of practices. The first ANOVA yielded a significant effect for type of practice, $F(3, 486) = 93.10$, $p < .001$, $\eta_p^2 = .37$. In agreement with expectations and the findings of Study 1, pairwise comparison (LSD) indicated that the level of tolerance was highest for the wearing of a headscarf ($M = 2.33$, $SD = 0.98$) and lowest for the imam's speech in which homosexuals are called inferior people ($M = 1.99$, $SD = 0.85$). In between these two were the social-conventional cases of the schools ($M = 2.16$, $SD = 0.94$) and the shaking of hands

($M = 2.20$, $SD = 0.95$). No significant differences between these latter two practices were found. Thus, tolerance was highest for the practice generally perceived as a personal decision, followed by the predominantly nonconventional practices, and lowest for the practice that was most clearly considered to be morally wrong. However, the second ANOVA revealed that the overall pattern was qualified by significant interactions with educational level, $F(6, 2856) = 3.92$, $p = .001$, $\eta_p^2 = .01$; gender, $F(3, 2856) = 13.06$, $p < .001$, $\eta_p^2 = .01$; and age, $F(3, 2856) = 2.75$, $p < .05$, $\eta_p^2 = .003$. These variables had significant main effects as well, $F(2, 952) = 7.23$, $p < .001$, $\eta_p^2 = .02$; $F(1, 952) = 11.18$, $p < .001$, $\eta_p^2 = .01$; and $F(1, 952) = 16.90$, $p < .001$, $\eta_p^2 = .02$, respectively.

Results for age indicated that older adolescents were generally less tolerant than younger adolescents. However, further inspection of each separate practice showed that the negative linear effect of age was strongest in case of the schools and nonsignificant in case of the handshaking. For ease of presentation, means are presented for three different age groups in Table 3.

As expected, for education we found that students in the highest level of education were more tolerant of Muslim practices than students in the lowest level (Table 3). However, post hoc analyses showed that for the imam's speech no difference in acceptance existed between the educational levels. Finally, there were differences between boys and girls for all cases, but the direction of the relation between gender and tolerance was not the same for each case (Table 3). Girls were more tolerant than boys with the exception of the imam's speech. For this particular case, boys were significantly more tolerant than girls.

Tolerance of persuasion. The first ANOVA with tolerance of persuading fellow Muslims as the dependent repeated measures variable yielded a significant effect for the type of practice, $F(3, 2898) = 162.32$, $p < .001$, $\eta_p^2 = .16$. Pairwise comparison showed, again, that the least tolerant judgments were given with respect to the imam persuading others ($M = 1.81$, $SD = 0.95$), followed by the female Muslim teacher ($M = 2.08$, $SD = 0.99$), and the students wearing a headscarf ($M = 2.34$, $SD = 1.10$). Unexpectedly, however, the level of tolerance was highest with respect to the Islamic schools ($M = 2.52$, $SD = 1.09$).

The second ANOVA showed that there was a significant main effect of age, $F(1, 960) = 4.33$, $p < .05$, $\eta_p^2 = .004$, and significant interactions between type of practice, and educational level and gender, respectively, $F(6, 2880) = 2.85$, $p < .01$, $\eta_p^2 = .01$, and $F(3, 2880) = 5.65$, $p < .001$, $\eta_p^2 = .01$. Table 4 shows mean levels of tolerance by educational level, gender, and age. Again, and as expected, older participants were less likely to give tolerant judgments. Next, educational level had only an effect for the case of the imam's speech: The higher educated were even less tolerant of the imam trying to persuade other Muslims. Boys and girls differed from each other in their level of tolerance of a Muslim teacher persuading others not to shake hands with men, and in the case of the imam. In the former, girls were found to be more tolerant than boys, in the latter it was the other way around.

Table 3
Mean Scores and Standard Deviations (in Parentheses) for Tolerance of the Four Types of Behavior by Educational Level, Gender, and Age: Study 2

	Headscarf	School	Hand	Speech	Total
Educational level					
Low	2.16 (1.01)	2.12 (1.03)	2.11 (0.98)	2.01 (0.94)	2.10 (0.70)
Middle	2.32 (0.96)	2.11 (0.88)	2.20 (0.91)	1.96 (0.86)	2.15 (0.62)
High	2.56 (0.92)	2.26 (0.90)	2.31 (0.89)	1.98 (0.73)	2.27 (0.57)
Gender					
Boy	2.23 (0.97)	2.03 (0.99)	2.08 (0.95)	2.05 (0.97)	2.10 (0.67)
Girl	2.44 (0.99)	2.29 (0.87)	2.32 (0.91)	1.91 (0.70)	2.24 (0.60)
Age (years)					
13–14	2.42 (0.99)	2.35 (0.97)	2.23 (0.94)	2.09 (0.92)	2.29 (0.66)
15	2.39 (0.99)	2.18 (0.94)	2.22 (0.91)	1.97 (0.83)	2.19 (0.65)
16–17	2.18 (0.94)	1.98 (0.90)	2.14 (0.95)	1.90 (0.82)	2.05 (0.60)

Note. Four-point scale (1–4).

Table 4
Mean Scores and Standard Deviations (in Parentheses) for Tolerance of the Persuasion of Other Muslims by Educational Level, Gender, and Age

	Headscarf	School	Hand	Speech	Total
Educational level					
Low	2.32 (1.17)	2.48 (1.11)	2.08 (1.02)	1.92 (0.98)	2.20 (0.80)
Middle	2.30 (1.04)	2.53 (1.12)	2.10 (1.00)	1.77 (0.95)	2.17 (0.82)
High	2.42 (1.07)	2.56 (1.01)	2.05 (0.93)	1.72 (0.92)	2.19 (0.77)
Gender					
Boy	2.32 (1.13)	2.48 (1.16)	2.01 (1.02)	1.88 (1.05)	2.17 (0.82)
Girl	2.37 (1.05)	2.56 (1.00)	2.14 (0.95)	1.73 (0.85)	2.20 (0.77)
Age (years)					
13–14	2.40 (1.09)	2.53 (1.06)	2.15 (1.04)	1.88 (0.92)	2.24 (0.77)
15	2.37 (1.08)	2.57 (1.09)	2.07 (0.95)	1.78 (0.97)	2.20 (0.80)
16–17	2.26 (1.21)	2.44 (1.10)	2.03 (1.00)	1.79 (0.96)	2.13 (0.81)

Note. Five-point scale (1–5).

Hence, these results are comparable to the previous analysis on the tolerance of the practice itself. It is important to note that, although differences exist for tolerance of the imam's speech between boys and girls, and between students in higher and lower levels of education, all participants tended to have an intolerant attitude toward this type of behavior. The scores lie well below the neutral midpoint (3) of the scale.

Tolerating Practices Versus Tolerating Persuasion

Next, we examined the difference between the two types of tolerance judgments. For tolerance of the practice itself, answer categories ranged from 1 to 4 with no neutral midpoint. In contrast, tolerance of the public expression in order to persuade others had a scale ranging from 1 to 5 with a neutral midpoint. Therefore, we rescaled the items of both types of questions with the following formula: (item − 1)/(number of answer categories − 1). In addition, we examined the frequencies of the answers given in order to draw meaningful conclusions.

We tested the difference in the two types of tolerance judgments by performing for each of the four cases a simple repeated measures ANOVA, with the type of tolerance as the repeated measure. For the case about the school, the two tolerance scores did not significantly differ. However, and as expected, for the other practices the level of tolerance of the public expression was lower than for the practice itself ($ps < .001$). Additional analyses showed that this pattern was consistent across age, educational level, and gender. Furthermore, for all four cases there were clear differences in percentages of tolerant versus intolerant responses. For the wearing of a headscarf, we found that 43% of the participants thought that this should be accepted. However, trying to persuade other Muslims to wear a headscarf was considered acceptable by only 15% (with an additional 26% answering "neutral"). For the other three cases, we obtained similar results. For the founding of Islamic schools, 35% of the participants thought that this should be accepted but only 18% found it acceptable that people try to persuade other Muslims to found Islamic schools (32% answers neutral). Furthermore, 43% thought that the decision of a female teacher not to shake hands with men should be accepted. However, for the persuasion of others, this percentage was only 10% (19% answers neutral). The percentages for the imam's speech were respectively 21% and 6% (neutral: 15%). These percentages indicate that the participants were less tolerant of attempts to persuade other Muslims than of the practices themselves.

In-Group Identification and Multiculturalism

In-group identification and the endorsement of multiculturalism were negatively associated, $r = -.40$, $p < .01$. We examined whether these two measures were associated with tolerance of the Muslims practices and of the campaigning for social support of fellow Muslims. For both types of tolerance judgments, we conducted a 4 (type of practice) × 2 (gender) × 3 (school level) repeated measures ANOVA, with type of practice as repeated measure, and with age, in-group identification, and multiculturalism as covariates.

Tolerance of practices. For tolerance of the Muslim practices, the between-subjects results indicate a significant effect for the continuous variable multiculturalism, $F(1, 950) = 225.20$, $p < .001$, $\eta_p^2 = .19$. As hypothesized, students who more strongly endorsed a multicultural ideology were more tolerant of Muslim practices. This effect was qualified by an interaction with the type of practice $F(3, 2868) = 24.10$, $p < .001$, $\eta_p^2 = .03$. Simple main effects analyses showed that although multiculturalism was significantly associated with tolerance for all practices, the strength of the relationship differed between the practices. As expected, estimates of the effect size indicated that the effect was much weaker for the imam's speech ($\eta_p^2 = .03$) and strongest in the case of the headscarf ($\eta_p^2 = .13$). The cases of the schools ($\eta_p^2 = .10$) and the shaking of hands ($\eta_p^2 = .11$) were in between.

We also examined the extent to which students endorsing high versus low levels of multiculturalism differentiated between the four different practices. That is to say, we compared students with scale scores equal or above the median (2.75) to students scoring below the median. Means and standard deviations for both groups are given in Table 5. Post hoc tests (LSD) showed that participants who scored high on multiculturalism showed the overall differentiation pattern (highest tolerance for headscarf, lowest tolerance for the imam's speech, with tolerance for schools and handshaking in between). Yet, participants scoring low on multiculturalism only differentiated between the case of the headscarf on the one hand, and the school, handshaking, and speech cases on the other hand.

The repeated measures ANOVA did not reveal a main effect for in-group identification or an interaction effect with the type of practice. However,

Table 5

Mean Scores and Standard Deviations (in Parentheses) for Tolerance of Practices and Persuasion by Multiculturalism and Ethnic Identification (Median Splits)

	Headscarf	School	Hand	Speech
Tolerance of practices[a]				
Multiculturalism low	1.96 (0.89)	1.85 (0.93)	1.87 (0.88)	1.88 (0.89)
Multiculturalism high	2.70 (0.93)	2.45 (0.86)	2.52 (0.87)	2.08 (0.80)
Tolerance of persuasion[b]				
Multiculturalism low	2.09 (1.10)	2.26 (1.15)	1.84 (0.95)	1.73 (0.95)
Multiculturalism high	2.59 (1.03)	2.78 (0.96)	2.31 (0.97)	1.89 (0.96)

[a]Four-point scale (1–4). [b]Five-point scale (1–5).

given the considerable association between in-group identification and the endorsement of multiculturalism ($r = -.40$), we also explored the results of an ANOVA without the multiculturalism measure. This analysis yielded the expected main effect for identification, $F(1, 957) = 37.47$, $p < .001$, $\eta_p^2 = .04$, but no significant interaction effect with the type of practice, $p > .05$. Those identifying strongly as Dutch were less tolerant than those with a low level of identification.

Tolerance of persuasion. The same analyses were repeated with tolerance of the persuasion of fellow Muslims as the dependent variable. As expected, there were significant main effects for both multiculturalism, $F(1, 958) = 86.63$, $p < .001$, $\eta_p^2 = .08$, and for ethnic identification, $F(1, 960) = 3.89$, $p < .05$, $\eta_p^2 = .004$. Participants who endorsed multiculturalism relatively strong were more tolerant than those who were less in favor of multiculturalism. Furthermore, those identifying strongly as Dutch made less tolerant judgments compared to those with low levels of in-group identification.

As expected, the association between multiculturalism and tolerance was further qualified by an interaction effect with type of practice, $F(3, 2874) = 10.15$, $p < .001$, $\eta_p^2 = .01$.

Simple main effect analyses showed that the effect of multiculturalism was significantly smaller in the case of the imam's speech. Estimates of the effect size were comparable in the cases of the headscarf ($\eta_p^2 = .06$), schools ($\eta_p^2 = .06$), and shaking hands ($\eta_p^2 = .07$), and especially low with respect to the imam's speech ($\eta_p^2 = .01$). For in-group identification, no interaction with the type of behavior was found, also not when multiculturalism was excluded from the analyses.

Separate post hoc tests (LSD) for students with a weak versus a strong endorsement of multiculturalism (median split; see Table 5) showed that both groups had the same differentiation pattern with the highest tolerance for persuasion in case of the school, followed by the headscarf and the handshaking, and the lowest tolerance for the imam's speech. However, participants who endorsed multiculturalism relatively strongly made clearer differentiations than those who were less in favor of multiculturalism.

General Discussion

This research examined ethnically Dutch adolescents' tolerance of specific practices by Muslim actors. In trying to maximize the relevance and ecological validity of the research, we used realistic and currently debated issues instead of unfamiliar and fictitious scenarios. In culturally diverse societies, controversies and conflicts are typically about concrete practices and actions, and, at least in Europe, predominantly relate to Muslims (Modood & Ahmad, 2007; Zolberg & Long, 1999).

In general, we found that adolescents were rather negative toward Muslim actors and expressed moderate to low levels of tolerance. However, in agreement with social-cognitive domain theory and previous research findings (e.g., Verkuyten & Slooter, 2007, 2008), it turned out that the acceptance of different types of behavior depended on both content and context. Using domain theory (Turiel, 1983), we expected that the tolerance of Muslim practices depends on the domain that the practices predominantly evoke. According to the theory, people make a distinction between the personal, the conventional, and the moral domains but what content-wise is considered to belong to each of these domains is not self-evident, particularly not in relation to the personal and the social-conventional. Therefore, in Study 1 we first examined whether Dutch adolescents make an empirical distinction between the three domains for the different Muslim practices. We presented the participants with four different cases: Muslim students wearing a headscarf, parents founding separate Islamic schools, a female Muslim teacher refusing to shake hands with men, and an imam who publicly calls homosexuals inferior people. The findings of Study 1 indicate that although these cases are multifaceted, they do differ in the salience of the particular domain they evoked, with the first case predominantly involving the personal domain, the second and third triggering more social-conventional concerns, and the fourth raising moral issues. The distinction between the three domains was found independently of age, educational level, and gender.

Study 2 examined two types of tolerance judgments: tolerance of the particular practices and of campaigns for public support of co-Muslims for these practices. In agreement with the domain theory, it turned out that tolerance was highest in the personal domain and lowest in the moral domain, with the social-conventional domain in between. Thus, the participants were found to be most tolerant of Muslim students wearing a headscarf and least tolerant of the imam's speech. The cases of the Islamic school and the refusal to shake hands, both mainly social-conventional matters, fell in between. This difference was found for both types of tolerance: for the actual practices and for Muslims trying to persuade others to engage in the same practices. The one exception was that campaigning for separate Islamic schools was accepted more than trying to persuade other Muslims to wear a headscarf.

There was a clear difference in both types of tolerance. Previous studies have found that people tend to be more tolerant of campaigns for public support for a dissenting belief than of the practice based on that belief (e.g., Verkuyten & Slooter, 2007; Wainryb et al., 1998; Wainryb et al., 2001). Not only is the higher acceptance of the public expression of dissenting beliefs consistent with the idea of free speech, it can be seen as stimulating debate and as causing less direct harm or injustice than the actual act. However, the current findings clearly show that the adolescents were less tolerant of persuading fellow Muslims than of the practice itself. Campaigning for support and persuading others implies mobilizing Muslims, for example, to start wearing a headscarf, to stop shaking hands with people of the opposite sex, and to found more Islamic schools. Politicians and the media tend to present these practices as "backward" and as threatening Dutch identity and culture (see Scroggins, 2005; Vasta, 2007; Verkuyten & Zaremba, 2005). They would undermine the secular and Christian traditions of the Netherlands. Trying to persuade fellow Muslims to act similarly is likely to be seen as contributing to the "Islamization of Dutch society" and therefore leads to lower acceptance compared to the act itself. Perceived threat is considered the most important predictor of intolerance (see Gibson, 2006), but it should be noted that we did not measure feelings of threat. However, in a previous study it was found that these feelings underlie Dutch adolescents' intolerance of Muslims (Van der Noll et al., 2010).

The importance of intergroup factors for understanding tolerance judgments was also found in relation to in-group identification and the endorsement of multiculturalism. Using SIT (Tajfel & Turner, 1979), we predicted that participants identifying strongly as Dutch would be less tolerant. This turned out to be the case for the practice itself (in the analysis without multiculturalism) and for the persuasion of fellow Muslims to act similarly. However, the findings for the endorsement of multiculturalism were stronger than for in-group identification and a similar result has been found in another study on Dutch adolescents' tolerance of Muslims (Van der Noll et al., 2010). One reason is that in-group identification is typically associated with in-group evaluation and not necessarily with the evaluation of out-groups, whereas, in the European context, multiculturalism is predominantly seen as focusing on out-groups (Verkuyten, 2006). In addition, compared to group identification, multicultural ideology is more closely related to questions of how to deal with dissenting beliefs and practices. Thus, multiculturalism seems to provide a general ideological view about the importance of cultural diversity that emphasizes that people should be recognized and valued in their group identity, and that there should be social equality and equal opportunities. This result is in agreement with research that has shown that beliefs about democratic processes and the protection of minority rights is a primary source of political tolerance (McClosky & Brill, 1983; Sullivan & Transue, 1999).

However, these results indicate that adolescents' tolerance judgments of out-group behaviors are driven not only by intergroup factors but also by moral concerns. Participants high on multiculturalism tended to tolerate practices that raise mainly personal and conventional considerations but were less accepting of the practice that was predominantly perceived as a moral transgression. The endorsement of the value of cultural diversity appears not to be unlimited. Multiculturalism does not imply moral relativism in which all practices and ideas are judged equally right and acceptable (Lukes, 2008). Respecting other cultural communities and their practices is difficult when concerns and considerations of others' welfare and fairness are involved. Already at the age of 5, children have been shown to think that some beliefs are relative and others not, and that their judgments of relativism differ from their tolerant judgments of divergent beliefs (Wainryb, Shaw, Langley, Cottam, & Lewis, 2004).

We also examined the development of tolerance by focusing on the effects of age, education, and gender. In contrast to the idea of an age-related progression from less to more principled reasoning and thus a more tolerant attitude (e.g. Enright & Lapsley, 1981; Enright et al., 1984), our results actually show the opposite. Older participants were found to be less tolerant across the range of practices and of both the practice itself and the persuasion of others. These results are in line with findings of a previous study that found among Dutch adolescents a negative association between age and the acceptance of Muslims founding separate schools (Verkuyten & Slooter, 2007). As argued by Killen and Stangor (2001), with age children become increasingly concerned about the nature of social groups and about group norms and expectations (Abrams & Rutland, 2008). It is thus likely that the older participants are more concerned about the functioning of the Dutch society and tend to see the Muslim practices as threatening the norms and values of their in-group. Hence, similar to research that has examined tolerance and children's and adolescents' reasoning about civil liberties (e.g., Helwig, 1995, 1997; Ruck et al., 1998; Wainryb et al., 1998; Wainryb et al., 2001), the current findings do not support a global stage interpretation. Rather, the findings suggest that with age adolescents tend to evaluate concepts of rights and freedoms increasingly in relation to other considerations and concerns, including the intergroup context.

In agreement with previous findings (see Verkuyten & Slooter, 2007; Vogt, 1997), the level of education of the participants was significantly and positively associated with tolerance. However, this effect differed between the types of practices that the participants were asked to evaluate (Wainryb et al., 1998). Educational differences were largest in the case of the headscarf and almost absent in relation to the moral issue of the imam's speech. Compared to participants in the lower levels of education those in the higher levels differentiated more between the different types of practices. Chong (1993) found that in discussing group rights, higher educated students tended to focus on the right itself, whereas the lower educated tended to focus on the group. For our research, this might mean that adolescents in the lowest level of education were predominantly concerned about the outgroup of Muslims whereas the higher educated tended to evaluate the nature of the practices. This interpretation implies that a focus on the actual practices rather than on the group itself does not lead to higher tolerance of persuading fellow Muslims, on the contrary.

Girls were found to be more tolerant than boys in the cases of the headscarf, the school, and the shaking of hands. However, for the imam's speech the picture was reversed, with girls being somewhat less tolerant than boys. It could be argued that females are more likely to reason from a care or welfare perspective (Gilligan, 1982), which makes them more sensitive to the potential (psychological) harm that is inflicted upon homosexuals by the imam's speech. However, systematic research has shown that gender does not explain variability in moral reasoning development (see Walker, 2006). Furthermore, the care perspective does not explain why female adolescents were actually more tolerant than male adolescents in the three other cases that involved freedoms and rights. In two of these three cases, the actors were women and this could have made the female participants' gender identity salient. As a result, their responses might have been influenced by their shared gender identity with the Muslim females in the scenarios. Furthermore, children know that males are considered of higher status than females and the female participants' higher tolerance for these two practices might be a reaction to the lower status and power of women (see Powlishta, 2004). It is also possible that females are more sensitive to issues of rights and freedoms because they have more experiences with gender discrimination and exclusion (e.g., Theimer, Killen, & Stangor, 2001). Future studies could test these different interpretations, for example, by presenting the participants with various practices that were carried out by both male and female Muslims.

Questions on the acceptance of minority practices and minority rights can involve many different situations and issues, and tolerance might also depend on the national context. Islam has moved to the center of debates in most European countries and the retreat of multicultural policies is not restricted to the Netherlands (Joppke, 2004; Zolberg & Long, 1999). Muslims are typically perceived as threatening and they are evaluated rather negatively by majority group populations across Europe (Strabac & Listhaug, 2008). There are strong debates about the acceptance and accommodation of Islamic practices within most liberal states. So, it is likely that our findings apply to other national contexts. Perhaps, the mean level of tolerance for the various practices differs across contexts, but the differentiation between the different domains, as proposed by sociocognitive domain theory, and the role of the endorsement of multiculturalism and in-group identification, as proposed by intergroup approaches, are likely to be similar. Future studies should examine this generality issue more closely and these studies could also examine the judgments of Muslim adolescents (Verkuyten & Slooter, 2008).

In this research, we used questions with precoded answer categories and we did not ask participants to give their reasons for tolerating or not tolerating particular practices. But children and adolescents form judgments on the basis of various concerns and considerations that are evaluated and assessed in relation to each other and the situation to which these apply (Turiel, 2002). That means that there is flexibility of thought and that the same practice can raise very different concerns. For example, the wearing of a headscarf does not only have to be considered a personal choice but can also raise social-conventional concerns and moral issues of gender equality. Similarly, the practice of founding Muslim schools can raise important conventional concerns but typically also involves matters of equal rights. But the fact that different interpretations and justifications are possible does not mean that socially shared or dominant interpretations do not exist. Both of our studies show that, in general and from the perspective of adolescents, some practices more than others raise moral, or conventional, or rather personal concerns. Calling homosexuals inferior people is typically considered harmful and rejected by the great majority of our participants, whereas the wearing of a headscarf is seen more as a personal decision. Thus, we focused on the distinctions made by adolescents themselves and we examined their tolerance judgments in relation to this. However, it would be useful to study adolescents' social reasoning about these topics and to examine the ways in which they weigh and coordinate various considerations in making judgments.

In conclusion, we drew on social-cognitive domain theory and SIT for understanding adolescents' tolerance of practices by Muslims actors, and we demonstrated the fruitful outcome for combining these theoretical frameworks. In studying out-group evaluations and behaviors, an increasing number of studies use an integrated framework by drawing on both developmental and social-psychological theories (see Killen et al., 2006). Previous research, for example, has studied the role of group identity and stereotypic expectations in children's reasoning about peer inclusion and exclusion (e.g., Abrams et al., 2008; Killen & Stangor, 2001). We have tried to make a contribution to this literature by focusing on aspects of tolerance regarding concrete Muslim practices that are, or recently were, debated in the Dutch context. Adolescents were found to make a distinction between practices that they evaluated negatively but regarded mainly a personal matter, practices that predominantly violate conventional norms, and practices that violate moral norms. The level of tolerance differs between these domains and intergroup factors appear to be less important for the acceptance of out-group practices that evoke moral concerns. Our findings indicate the importance of using an integrative social-developmental perspective. By combining cognitive domain theory and SIT, we were able to formulate specific predictions and improve our understanding of adolescents' tolerance of dissenting practices.

This understanding might be of use for developing adequate interventions to improve tolerance, which is foundational for equality and the development of harmonious intergroup relations. Most lines of thinking argue that the reduction of stereotypes and prejudice is necessary for these kinds of relationships to develop. However, our knowledge and ability to reduce stereotypes and prejudice remains limited. Generalized perceptions and negative beliefs and feelings do not appear to be easy to change or to reject. The importance of tolerance is that it keeps these beliefs and feelings from becoming negative actions, thereby forming the first crucial step toward civility or the last barrier to conflict (Vogt, 1997). Tolerance also does not imply relativism in which it is argued that one should refrain from value judgments in assessing other groups (Lukes, 2008). Tolerance always has limits and does not imply a full acceptance and valuing of all social practices of other groups, such as potentially harmful activities, illiberal internal rules, and inequalities. A decision of whether a particular practice should be tolerated always involves a variety of considerations and the results show that adolescents evaluate different practices differently. This means, for example, that effective civics education has to focus on the related questions of what should and what should not be tolerated and why. Furthermore, our findings suggest that an educational emphasis on cultural diversity and multicultural recognition is a promising avenue for improving majority group adolescents' tolerance of Muslims. Participants who endorse multiculturalism more strongly appear to be more tolerant, but much less so when it comes to the imam's harmful speech. Thus, it seems important to stimulate multicultural recognition but without the relativism that is found in some forms of multiculturalism that tend to celebrate diversity unconditionally.

References

Abrams, D., & Rutland, A. (2008). The development of subjective group dynamics. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 47–65). Oxford, UK: Oxford University Press.

Abrams, D., Rutland, A., Ferrell, J. M., & Pelletier, J. (2008). Children's judgments of disloyal and immoral peer behavior: Subjective group dynamics in minimal intergroup contexts. *Child Development*, *79*, 444–461.

Arends-Tóth, J. V., & van de Vijver, F. J. R. (2003). Multiculturalism and acculturation: Views of Dutch and Turkish-Dutch. *European Journal of Social Psychology*, *33*, 249–266.

Berry, J. W. (2006). Mutual attitudes among immigrants and ethnocultural groups in Canada. *International Journal of Intercultural Relations*, *30*, 719–734.

Berry, J. W., & Kalin, R. (1995). Multicultural and ethnic attitudes in Canada. *Canadian Journal of Behavioral Science*, *27*, 301–320.

Bizman, A., & Yinon, Y. (2001). Intergroup and interpersonal threats as determinants of prejudice: The moderating role of in-group identification. *Basic and Applied Psychology*, *23*, 191–196.

Bobo, L. D., & Licardi, F. C. (1989). Education and political tolerance. *Public Opinion Quarterly*, *53*, 285–308.

Chong, D. (1993). How people think, reason, and feel about rights and liberties. *American Journal of Political Science*, *37*, 867–899.

Enright, R. D., & Lapsley, D. K. (1981). Judging others who hold opposite beliefs: The development of belief-discrepancy reasoning. *Child Development*, *52*, 1053–1063.

Enright, R. D., Lapsley, D. K., Franklin, C. C., & Steuck, K. (1984). Longitudinal and cross-cultural validation of the belief-discrepancy reasoning construct. *Developmental Psychology*, *20*, 143–149.

Gibson, J. L. (2006). Enigmas of intolerance: Fifty years after Stouffer's communism, conformity, and civil liberties. *Perspectives on Politics*, *4*, 21–34.

Gijsberts, M. (2005). Opvattingen van autochtonen en allochtonen over de multi-etnische samenleving. In SCP, WODC, & CBS (Eds.), *Jaarrapport Integratie 2005* (pp. 189–205). The Hague, Netherlands: SCP, WODC, and CBS.

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Helwig, C. C. (1995). Adolescents' and young adults' conceptions of civil liberties: Freedom of speech and religion. *Child Development*, *66*, 152–166.

Helwig, C. C. (1997). The role of agent and social context in judgments of freedom of speech and religion. *Child Development*, *68*, 484–495.

Joppke, C. (2004). The retreat of multiculturalism in the liberal state: Theory and policy. *British Journal of Sociology*, *55*, 237–257.

Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. T. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117–131.

Killen, M., Margie, N. G., & Sinno, S. (2006). Morality in the context of intergroup relations. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 155–183). Mahwah, NJ: Erlbaum.

Killen, M., & McKown, C. (2005). How integrative approaches to intergroup attitudes advance the field. *Applied Developmental Psychology*, *26*, 616–622.

Killen, M., Sinno, S., & Margie, N. G. (2007). Children's experiences and judgments about group exclusion and inclusion. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 35, pp. 173–218). New York: Elsevier.

Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development*, *72*, 174–186.

Kim, J. M., & Turiel, E. (1996). Korean and American children's concepts of adult and peer authority. *Social Development*, *5*, 310–329.

Laupa, M., & Turiel, E. (1993). Children's concepts of authority in social contexts. *Journal of Educational Psychology*, *85*, 191–197.

Levy, S. R., West, T. L., & Ramirez, L. (2005). Lay theories and intergroup relations: A social-developmental perspective. *European Review of Social Psychology*, *16*, 189–220.

Lukes, S. (2008). *Moral relativism*. London: Profile Books.

McClosky, H., & Brill, A. (1983). *Dimensions of tolerance: What Americans believe about civil liberties*. New York: Basic Books.

Modood, T., & Ahmad, F. (2007). British Muslim perspectives on multiculturalism. *Theory, Culture & Society*, *24*, 187–213.

Nesdale, D. (2008). Peer group rejection and children's intergroup prejudice. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 32–46). New York: Oxford University Press.

Nucci, L. P. (2001). *Education in the moral domain*. Cambridge, UK: Cambridge University Press.

Pew Research Center. (2005). Islamic extremism: Common concern for Muslim and Western publics [online]. The Pew Global Attitudes Project. Retrieved November 23, 2008, from http://pewglobal.org/files/pdf/248.pdf

Pomerantz, E. M., & Newman, L. S. (2000). Looking in on the children: Using developmental psychology as a tool for hypothesis testing and model building in social psychology. *Personality and Social Psychology Review*, *4*, 300–316.

Powlishta, K. K. (2004). Gender as a social category: Intergroup processes and gender-role development. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 103–134). New York: Psychology Press.

Ruck, M. D., Abramovitch, R., & Keating, D. P. (1998). Children's and adolescents' understanding of rights: Balancing nurturance and self-determination. *Child Development*, *64*, 404–417.

Scroggins, D. (2005). The Dutch-Muslim culture war. *The Nation*, June 27, 21–25.

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb

(Eds.), *The emergence of morality in young children* (pp. 1–82). Chicago: University of Chicago Press.

Sigelman, C. K., & Toebben, J. L. (1992). Tolerant reactions to advocates of disagreeable ideas in childhood and adolescence. *Merrill-Palmer Quarterly, 38*, 542–557.

Smetana, J. G. (1995). Morality in context: Abstractions, ambiguities, and applications. In R. Vasta (Ed.), *Annals of child development* (Vol. 10, pp. 83–110). London: Jessica Kingsley.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 117–153). Mahwah, NJ: Erlbaum.

Sniderman, P., & Hagendoorn, L. (2007). *When ways of life collide: Multiculturalism and its discontents in the Netherlands*. Princeton, NJ: Princeton University Press.

Strabac, Z., & Listhaug, O. (2008). Anti-muslim prejudice in Europe: A multilevel analysis of survey data from 30 countries. *Social Science Research, 37*, 268–286.

Sullivan, J. L., & Transue, J. E. (1999). The psychological underpinnings of democracy: A selective review of research on political tolerance, interpersonal trust and social capital. *Annual Review of Psychology, 50*, 625–650.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Tausch, N., Tam, T., Hewstone, M., Kentworthy, J., & Cairns, E. (2007). Individual-level and group-level mediators of contact effects in Northern Ireland: The moderating role of social identification. *British Journal of Social Psychology, 46*, 541–556.

Theimer, C. E., Killen, M., & Stangor, C. (2001). Young children's evaluations of exclusion in gender-stereotypic peer contexts. *Developmental Psychology, 37*, 18–27.

Tisak, M. S., Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology, 24*, 352–357.

Turiel, E. (1983). *The development of social knowledge. Morality and convention*. Cambridge, UK: Cambridge University Press.

Turiel, E. (1998). The development of morality. In N. Eisenberg (Vol. Ed.) & W. Damon (Series Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–932). New York: Wiley.

Turiel, E. (2002). *The culture of morality*. Cambridge, UK: Cambridge University Press.

Van der Noll, J., Verkuyten, M., & Poppe, E. (2010). Political tolerance and prejudice: Differential reactions towards Muslims in the Netherlands. *Basic and Applied Social Psychology, 32*, 46–56.

Vasta, E. (2007). From ethnic minorities to ethnic majority policy: Multiculturalism and the shift to assimilation in the Netherlands. *Ethnic and Racial Studies, 30*, 713–740.

Velasco Gonzalez, K., Verkuyten, M., Weesie, J., & Poppe, E. (2008). Prejudice towards Muslims in the Netherlands: Testing integrated threat theory. *British Journal of Social Psychology, 47*, 667–685.

Verkuyten, M. (2005). Ethnic group identification, and group evaluations among minority and majority groups: Testing the multiculturalism hypothesis. *Journal of Personality and Social Psychology, 88*, 121–138.

Verkuyten, M. (2006). Multicultural recognition and ethnic minority rights: A social identity perspective. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 17, pp. 148–184). London: Psychology Press.

Verkuyten, M., & Slooter, L. (2007). Tolerance of Muslim beliefs and practices: Age related differences and context effects. *International Journal of Behavioral Development, 31*, 467–477.

Verkuyten, M., & Slooter, L. (2008). Muslim and non-Muslim adolescents' reasoning about freedom of speech and minority rights. *Child Development, 79*, 514–528.

Verkuyten, M., & Zaremba, K. (2005). Inter-ethnic relations in a changing political context. *Social Psychology Quarterly, 68*, 375–386.

Vogt, W. P. (1997). *Tolerance and education: Learning to live with diversity and difference*. London: Sage.

Wainryb, C. (1993). The application of moral judgments to other cultures: Relativism and universality. *Child Development, 64*, 924–933.

Wainryb, C. (2006). Moral development in culture: Diversity, tolerance, and justice. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 211–240). Mahwah, NJ: Erlbaum.

Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development, 75*, 687–703.

Wainryb, C., Shaw, L. A., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreement. *Developmental Psychology, 37*, 373–386.

Wainryb, C., Shaw, L. A., & Maianu, C. (1998). Tolerance and intolerance: Children's and adolescents' judgments of dissenting beliefs, speech, persons, and conduct. *Child Development, 69*, 1541–1555.

Walker, L. J. (2006). Gender and morality. In K. Killen & J. Smetena (Eds.), *Handbook of moral development* (pp. 93–115). Mahwah, NJ: Erlbaum.

Witenberg, R. T. (2002). Reflective racial tolerance and its development in children, adolescents and young adults: Age related difference and context effects. *Journal of Research in Education, 12*, 1–8.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology, 78*, 635–664.

Zolberg, A., & Long, L. W. (1999). Why Islam is like Spanish: Cultural incorporation in Europe and the United States. *Politics and Society, 27*, 5–38.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1400–1416

The Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study

Elizabeth Pellicano

Institute of Education, London, and University of Western Australia

This longitudinal study tested the veracity of one candidate multiple-deficits account of autism by assessing 37 children with autism (*M* age = 67.9 months) and 31 typical children (*M* age = 65.2 months) on tasks tapping components of theory of mind (ToM), executive function (EF), and central coherence (CC) at intake and again 3 years later. As a group, children with autism showed poor false-belief attribution, planning ability, and set-shifting, together with enhanced local processing at both time points. At an individual level, however, the profile was far from universal at either intake or follow-up. Moreover, autistic children demonstrated significant changes over time in ToM and EF, but not CC, over the 3-year period. The challenges these findings pose for a multiple-deficits account are discussed.

The past few decades have seen considerable research efforts devoted to isolating a single underlying cognitive deficit that could provide a unifying explanation for the ''triad of impairments'' in autism—the severe difficulties in reciprocal social interaction and communication, and restricted interests and repetitive behaviors (American Psychiatric Association [APA], 1994). Three cognitive theories in particular have dominated the field: (a) the theory-of-mind (ToM) hypothesis, which claimed that autism is caused primarily by a specific inability to impute mental states to oneself and to others (Baron-Cohen, Leslie, & Frith, 1985); (b) the executive dysfunction hypothesis, which proposed that the symptoms of autism are a result of a primary problem in the planning and execution of complex actions (Hughes & Russell, 1993; Ozonoff, Pennington, & Rogers, 1991); and (c) ''weak'' central coherence (CC) theory, which posited that inherent to autism is an unusual tendency to focus on individual, local elements rather than global wholes (Frith, 1989; Happé & Frith, 2006).

The notion of a single cause at the cognitive level of analysis has, however, been challenged on two main grounds. First, all three influential ''single-deficit'' models have struggled to satisfy key criteria for an explanatory account of autism, namely, that the putative deficit should be ''universal, specific, and necessary and sufficient to cause the symptoms of the disorder'' (Pennington & Ozonoff, 1996, p. 57), and show persistence or stability over time (Ozonoff & McEvoy, 1994; Rutter, 1983; see Rajendran & Mitchell, 2007, for a review). Second, recent research has called into question the very premise underpinning the search for a core cognitive deficit: that the behaviors central to autism cluster together more often that what would be expected by chance as a result of shared etiology (Happé & Ronald, 2008; Happé, Ronald, & Plomin, 2006; Mandy & Skuse, 2008). Investigations of the broader autism phenotype have demonstrated that among those relatives of individuals with autism who display subclinical autistic features, the majority of these family members show behaviors in isolated symptom domains (e.g., social difficulties in the absence of repetitive behaviors, etc.) rather than a combination of these symptoms, albeit in more subtle form than in autism itself (e.g., Bolton et al., 1994). Furthermore, recent findings from a population-based study investigating autism-related traits in over three thousand 7- to 9-year-old typically developing twin pairs found that the cross-trait genetic correlations were surprisingly modest to low both across the general population and in children lying at the extreme end of the distributions, suggesting that largely independent genes may be

---

This research was supported by the British Academy, the Experimental Psychology Society (UK), and University College, Oxford. The author is grateful to all the families for their continued support, and to Murray Maybery and Marc Stears for helpful discussions.

Correspondence concerning this article should be addressed to Elizabeth Pellicano, Department of Psychology and Human Development, Centre for Research in Autism and Education, Department of Psychology and Human Development, 25 Woburn Square, London WC1H 0AA, United Kingdom. Electronic mail may be sent to l.pellicano@ioe.ac.uk.

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0007

operating on each aspect of the triad of impairments in autism (Ronald et al., 2006). Also, a considerable number of children in this study displayed behavioral difficulties in isolation (e.g., social difficulties only, communication problems alone). Both sets of findings provocatively suggest that the triad of impairments may in fact be dissociable, raising the possibility of distinct causes for each aspect of the behavioral phenotype.

In light of these challenges, Happé and colleagues have argued that ''at the cognitive level, as at the symptom/behavioral and genetic levels, autism may be characterized by fractionable impairments'' (Happé & Ronald, 2008, p. 296). Consequently, they have called for researchers to ''abandon the attempt to find a single cognitive explanation'' (Happé, Ronald, et al., 2006, p. 1219) in favor of an explanation encompassing coexisting atypicalities in *multiple* cognitive domains. These authors offered one candidate account in which a combination of cognitive atypicalities in three core domains—ToM, executive function (EF), and CC—exists in autism, where each atypicality underpins a distinct aspect of the behavioral phenotype. Since these atypicalities are held to be largely *independent* of each other at the phenotypic and genetic levels, autistic symptoms are therefore viewed as the result of multiple, *primary* cognitive atypicalities.

Recent evidence supports this view. Pellicano, Maybery, Durkin, and Maley (2006) found that young children with an autism spectrum disorder (ASD; $n = 40$) as a group did in fact demonstrate the cognitive profile proposed by Happé and colleagues, including difficulties in aspects of ToM (false-belief understanding) and EF (planning, cognitive flexibility) accompanied by weak CC (enhanced local information processing), relative to age- and ability-matched typical children. This support at the group level was somewhat tempered, however, by a failure to demonstrate the presence of this specific cognitive profile in each child with ASD.

One crucial question is how this specific cognitive profile manifests itself over time. Autism is generally understood as a pervasive, lifelong condition (APA, 1994), and therefore any putative explanation of the pathogenesis of autism should be able to account for the continuities and discontinuities that take place across the life span. Despite the wealth of research on specific cognitive skills in autism, however, there have been disappointingly few empirical studies tracking the *development* of such skills in autism—even within a single-deficit framework. The overarching goal of the current study therefore was to investigate longitudinally the multiple-deficits account proposed by Happé, Ronald, et al. (2006) focusing specifically on questions regarding developmental persistence—both at the group level and at the individual level.

Before describing this study it is worth considering briefly the findings of the five existing longitudinal studies. Remarkably, only four longitudinal investigations have traced developmental changes in ToM. Holroyd and Baron-Cohen (1993) followed 17 of 20 children with autism described originally in Baron-Cohen et al. (1985) and reassessed their false-belief understanding 7 years later. No overall gains in participants' ToM skills were found: Only 3 children passed the simple ''Sally–Ann'' test at follow-up (including 2 of 4 children who had succeeded on this task 7 years earlier), and no child passed a more advanced ToM test. Ozonoff and McEvoy (1994, Experiment 2) also reported no significant changes in the proportion of cognitively able autistic adolescents passing simple and more advanced ToM tasks over a 3-year period. Indeed, the scores of only 4 adolescents with autism improved over time. The authors therefore concluded that, overall, ToM abilities progress little with development.

Two recent studies, however, paint a more positive picture. Steele, Joseph, and Tager-Flusberg (2003) found significant improvements in autistic children's ToM skills over a 1-year period, with initial vocabulary level, rather than chronological age, predicting a significant (yet small; 3%) proportion of the variance in later ToM scores. Such gains, however, were driven largely by improvements on tasks tapping early-emerging ToM skills (understanding of pretense and desire) rather than on tasks of false-belief understanding. Serra, Loth, van Geert, Hurkens, and Minderaa (2002) also reported significant gains in a range of ToM skills over a 6-month period in preschool children with a mild variant of autism (pervasive developmental disorder–not otherwise specified; PDD–NOS), even though children with PDD–NOS developed at a slower rate than younger typically developing children. Their analysis, however, examined changes over time for an omnibus ToM score rather than for individual ToM task scores rendering it unclear whether such gains occurred for false-belief understanding per se.

To date, only two studies have investigated the developmental course of EF. In Ozonoff and McEvoy's (1994, Experiment 1) 3-year follow-up study, nonautistic adolescents with learning disability showed significant gains in planning efficiency (on the Tower of Hanoi) and made fewer perseverative responses (on the Wisconsin Card Sort Test [WCST]), but concomitant improvements were not evident in adolescents with autism. They concluded that, akin to their findings on ToM, difficulties in EF also persist over time. Griffith, Pennington, Wehner, and Rogers (1999) reported similar findings in preschoolers with autism on a spatial reversal task, a test of cognitive flexibility. Despite there being a trend for children with developmental delay to commit fewer perseverative errors at follow-up, autistic children's performance did not change significantly within the space of 1 year.

The findings from these longitudinal studies indicate that there are, on the whole, few changes in ToM or EF over time, implying that there might be a ceiling on the extent to which such abilities can develop in people with autism. Yet, there is wide variability in the age and ability of participants tested within and across the five studies, which not only makes comparisons between these few studies difficult, but also can potentially swamp longitudinal effects (i.e., improvements) within studies. Also, with the exception of Ozonoff and McEvoy's (1994) study, there has been no investigation of developmental changes in *multiple* cognitive domains; it is unclear, therefore, whether all core aspects of the cognitive phenotype (cf. Happé, Ronald, et al., 2006) take a similar developmental course. Indeed, no study has examined longitudinally local information processing in autism. The progression of CC therefore remains unknown.

The limited knowledge on the developmental course of each of the three cognitive domains presents an initial difficulty for a multiple-deficits account of autism. A complete picture of the development of the condition requires an understanding of the potential changes that take place over time within each domain, in addition to knowledge of the factors that influence such changes. This longitudinal study sought to take a crucial step in this direction. To this end, Pellicano et al.'s (2006) samples of cognitively able children with ASD and typically developing children were followed prospectively and reevaluated 3 years later on similar measures tapping key components of ToM, EF, and CC.

The primary aims of this study were twofold. The first aim was to test whether Happé, Ronald, et al.'s (2006) specific cognitive profile persisted across time both at the group level and at the individual level. If the putative cognitive profile—atypicalities in ToM, EF, and CC—is indeed central to autism, then the group of children with autism should show atypicalities in all three domains relative to a group of typically developing children both at intake and at follow-up (cf. Ozonoff & McEvoy, 1994). Furthermore, if the specific combination of atypicalities in ToM, EF, and CC underlies autistic symptoms as Happé et al. proposed, then this profile should be present in all, or almost all, autistic children at both time points.

The second aim of this study was to determine the nature and extent of developmental changes in ToM, EF, and CC, and to identify potential predictors of cognitive change. The majority of existing studies has failed to demonstrate developmental improvements in false-belief understanding and aspects of EF. It was expected, however, that longitudinal changes might be more readily detected in a sample of children with autism who were selected initially to be of at least average intellectual functioning and who fell within a restricted age range. Furthermore, language has been implicated in theoretical models of the development of both ToM (e.g., de Villiers, 2000) and EF (e.g., Russell, 1996), and has been linked empirically to autistic children's skills in ToM (e.g., Happé, 1995), EF (e.g., Liss et al., 2001; Pellicano, 2007), and CC (e.g., Pellicano et al., 2006). It was hypothesized, therefore, that for children with autism, individual differences in general cognitive ability, particularly verbal ability, would be influential in predicting developmental change in all three cognitive domains.

Method

Participants

A total of 90 children participated in the initial study (Time 1): 45 children with an ASD (40 boys) and 45 typically developing children (37 boys). Briefly, all children in the ASD group had received an independent clinical diagnosis of either autism ($n = 31$), PDD–NOS ($n = 12$), or Asperger syndrome ($n = 2$), according to *DSM–IV* criteria (APA, 1994) and met either full or partial criteria on the Autism Diagnostic Interview–Revised (ADI–R; Lord, Rutter, & Le Couteur, 1994; see Pellicano et al., 2006, for details). At intake, the ASD group had a mean age of 67.2 months ($SD = 10.51$, range = 49–88) and the typically developing group had a mean age of 65.1 months ($SD = 11.80$, range = 48–88). No child had a coexisting medical or developmental condition as reported by parents, or obtained a verbal or nonverbal IQ score below 80, as assessed by the Peabody Picture Vocabulary Test–Third Edition

(PPVT–III; Dunn & Dunn, 1997) and the Leiter International Performance Scale–Revised (Leiter–R; Roid & Miller, 1997), respectively. Note that unlike more standard IQ tests, which assess both receptive and expressive language ability, the measure used here to index verbal IQ assessed children's receptive vocabulary skills alone. Children with autism were recruited through community contacts in southwest Western Australia. The majority of children were White, and the parents were of mixed socioeconomic backgrounds, although specific data on socioeconomic status and educational attainment levels were not recorded.

Attempts were made to contact all families approximately 3 years after their involvement in the initial study to invite them to participate in the follow-up study (Time 2). Sixty-eight families were retraced and gave consent for reassessment (76% of the original sample; see Table 1). The mean time between the initial and follow-up studies was 32.9 months for the ASD group and 32.7 months for the typical group. Thirty-seven of 45 children with ASD (82%) were available for further testing. Of the 8 families who were unavailable for reassessment, 2 had relocated to a different state, while 6 families had moved and were untraceable. The 37 children with autism (33 boys) in the follow-up group were not significantly different from those who were not reassessed (*n* = 8) in terms of chronological age, *F*(1, 44) = 0.28, *p* = .60; verbal ability, *F*(1, 44) = 0.63, *p* = .43; nonverbal ability, *F*(1, 44) = 1.27, *p* = .27; or total algorithm score on the ADI–R, *F*(1, 44) = 1.90, *p* = .18. The follow-up group consisted of 26 children with autism, 9 children with PDD-NOS, and 2 children with Asperger syndrome, all of whom met full or partial criteria on the ADI–R conducted at Time 1 (social domain: *M* = 17.81, *SD* = 5.74; communication domain: *M* = 13.22, *SD* = 4.37; repetitive behaviors domain: *M* = 6.68, *SD* = 2.42). The Autism Diagnostic Observation Schedule–Generic (ADOS–G; Lord, Rutter, DiLavore, & Risi, 1999) was administered at follow-up to index these children's current sociocommunicative symptoms (Social Interaction score: *M* = 8.54, *SD* = 3.51; Communication score: *M* = 4.11, *SD* = 2.22; Total algorithm score: *M* = 12.62, *SD* = 4.98).

Table 1
Descriptive Statistics for Chronological Age, Verbal Ability, and Nonverbal Ability for the Autism Spectrum Disorder (ASD; n *= 37) and Typically Developing (*n *= 31) Groups at Both Time Points*

	ASD	Typical		
Variable	*M* (*SD*) Range	*M* (*SD*) Range	*F*	*p* value
Chronological age				
Time 1	67.92 (10.42) 49–88	65.19 (12.64) 48–88	0.95	.33
Time 2	100.84 (11.15) 80–123	97.94 (13.48) 75–122	0.94	.34
Verbal IQ (PPVT–III)				
Time 1	97.08 (11.52) 80–122	100.97 (8.72) 87–120	2.38	.13
Time 2	93.89 (17.88) 62–138	99.64 (10.53) 80–123	2.48	.12
Nonverbal IQ (Leiter–R)				
Time 1	113.27 (13.93) 83–141	115.61 (16.42) 89–147	0.40	.53
Time 2	104.35 (12.72) 80–135	106.97 (10.16) 93–129	0.85	.36

Note. PPVT–III = Peabody Picture Vocabulary Test–Third Edition; Leiter–R = Leiter International Performance Scale–Revised.

Of the initial sample of 45 typical children who were recruited from several local mainstream preschools and primary schools, 31 (69%) took part in the follow-up study (see Table 1). One family had relocated, 4 families declined to participate, and 10 families were untraceable. The 31 typical children (25 boys; 81%) involved in the follow-up study did not differ significantly from the 14 children who did not participate in terms of age, *F*(1, 44) = 0.02, *p* = .87; verbal ability, *F*(1, 44) = 1.66, *p* = .20; or nonverbal ability, *F*(1, 44) = 1.78, *p* = .19, at intake. Parents of typically developing children completed the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) to ensure that no child in this group displayed clinically significant levels of autistic symptomatology; all scores were well below the cutoff score of 15 for autism at Time 1 (*M* = 4.00, *SD* = 3.32) and Time 2 (*M* = 3.50, *SD* = 2.85).

Measures

Children completed a variety of measures assessing key aspects of cognition at each time point.

Verbal and nonverbal ability. At both time points, the PPVT–III (Dunn & Dunn, 1997) was used to assess children's receptive vocabulary and four subtests of the Leiter–R (Roid & Miller, 1997) were used to index nonverbal ability: *Matching* (a match-to-sample task using pictures of objects and abstract patterns), *Associated Pairs* (an associative memory task in which children were required to establish associations between pairs of pictured objects), *Forward Memory* (a visual short-term memory task that involved children copying the examiner's pointing sequence), and *Attention Sustained* (a visual attention task, which entailed identifying multiple target stimuli among distractor stimuli). Standard scores are reported in Table 1, but raw scores are used in correlational and regression analyses since such scores have not been adjusted for age, and therefore are more sensitive to developmental change.

Theory of mind. At each time point, children were presented with three standard false-belief prediction tasks to index ToM ability. For the *first-order unexpected-contents task* (based on Perner, Leekam, & Wimmer, 1987), children were shown three different boxes (e.g., milk carton) containing unexpected contents (e.g., rubber bands). After the lid of each box was replaced, they were asked to recall their own false belief (e.g., "Before you looked inside, what did you think was in the box?") and to predict another's false belief (e.g., "If I show this box to Mum, what will she think is inside?"). Children also answered a control question (e.g., "What is in the box really?") following each false-belief question. For the *first-order unexpected-transfer task* (based on Baron-Cohen et al., 1985), children were shown six different scenarios in which one character either displaced or substituted another character's object. At the end of each trial, children were asked a critical belief question (e.g., "Where will Sarah look for her apple?") and two control questions (e.g., "What is really in the bag?" and "What was in the bag in the beginning?"). For the *second-order unexpected-transfer task* (based on Perner & Wimmer, 1985), children were shown two displacement scenarios similar to those in the first-order unexpected-transfer task, although this time they witnessed the protagonist watching the transfer through a window. For each trial, they were asked to attribute a mistaken belief about a belief to a character (e.g., "Where will Tom think that Jane will look for her book?") as well as reality (e.g., "Where is the book really?") and memory (e.g., "Where did Jane put the book in the beginning?") control questions.

Children were given a score of 1 for each correctly answered false-belief test question. At both time points, children were asked 14 test questions: 6 points from the first-order unexpected-contents task (3 for recalling their own belief and 3 for recalling another's false belief), 6 points from the first-order unexpected-transfer task, and 2 points from the second-order unexpected-transfer task. All children correctly answered control questions, with the exception of 1 child with autism who failed the control questions on one second-order trial at Time 2. To avoid loss of data, this child was given a score of 0 of 2 on this task only.

Executive function. At both time points, children completed two measures each tapping a key component of EF, including planning ability and cognitive flexibility. Both EF components have been identified consistently as being disrupted specifically in autism (see Hill, 2004). The *Tower of London task* (Shallice, 1982; see also Hughes, 1998b) assessed children's higher order planning ability. Children were presented with three colored beads (red, white, black) arranged in a particular configuration (start state) on a wooden tower structure consisting of three vertical pegs of increasing size (small, medium, and big). They were then shown a picture of the beads in a different configuration (goal state) and asked to move the beads on their apparatus one at a time to match the goal state within the least possible number of moves (this number was clearly stated on the picture). After three practice trials, children were given problem sets of increasing difficulty (four trials of 1-, 2-, 3-, and 4-move problems). At Time 2, an additional 5-move problem set was included to increase task difficulty and extend the range of possible scores. Testing ceased if children failed all four trials within a problem set. The number of moves taken and rule violations were recorded. Children were given a score of 1 for each trial if they reached the goal state within the minimum number of moves and without violating any rules (Time 1: maximum score = 16; Time 2: maximum score = 20). High scores indicate good planning ability.

Cognitive flexibility was assessed using two developmentally sensitive card-sorting tasks. Both tasks were similar in nature to the traditional WCST (Heaton, 1981) and assessed the ability to switch flexibly between sorting categories in response to feedback. At Time 1, Hughes's (1998a) *teddy-bear set-shifting task* was used. There were three decks of cards, which differed in terms of either color (green vs. pink, blue vs. red, or yellow vs. purple), picture shown (hearts vs. diamonds, squares vs. moons, or stars vs. happy faces), and size of picture (small vs. large). To begin, children were presented with a teddy bear and one deck of cards, and were told that they were to work out which cards teddy liked best. If the card was one of teddy's favorites, the child posted it into a postbox. Alternatively, if the card was not one of teddy's favorites, then the child turned the card facedown on the table. Feedback was provided after each trial. When the child had successfully sorted six cards consecutively, or when a maximum of 20 trials had been presented, the sorting rule (e.g., color, shape, size) changed, upon which he or she was presented with a new teddy bear and new deck of cards. Following Hughes, children were not alerted to a change in sorting rule; this was implicit in the fact that children were presented with a new situation. Set-shifting performance was rated by the proportion of errors committed following the first sort to criterion.

At Time 2, children were administered a more difficult *computerized set-shifting task* (Comerford, 2005). Children were told that they would see one (target) card that appeared at the top of the screen, and that they would be required to match this card with one of four (response) cards presented at the bottom of the screen. There were 24 response cards, which varied on three dimensions (color, form, and background); all cards shared only one attribute with three of the four response cards and no attribute with the fourth card. Each card therefore could be sorted according to three rules (color, form, or background). Feedback was provided for each trial: The words *correct* or *incorrect* would appear in center screen. After the child had sorted six consecutive cards correctly, the rule changed (e.g., from color to form) without warning. Testing continued until the set of 24 cards had been presented twice (48 trials in total). To ensure compatibility with the earlier set-shifting task, the main variable of interest was the proportion of errors committed following the first sort to criterion. For both set-shifting tasks, a low score (i.e., minimal errors) indicates good cognitive flexibility.

Central coherence. Two visuospatial tasks were used to assess children's local processing bias: the *Children's Embedded Figures Test* (CEFT; Witkin, Oltman, Raskin, & Karp, 1971) and the *Pattern Construction task* from the Differential Ability Scales (Elliott, 1990). In the CEFT, children initially were shown a target shape (a triangle) and asked to find this shape hidden in a number of larger meaningful figures as quickly as possible (total 11 trials). There was a single target shape per figure, and children scored 1 point for each correctly identified target (maximum score = 11). The time taken to find the target also was recorded and was the dependent variable of interest. If children were unable to locate the target within 30 s, the maximum time (30 s) was given on that trial. Faster times are indicative of weak CC (i.e., better local processing).

In the Pattern Construction task, which is similar in nature to Wechsler's (1999) Block Design subtest, children were asked to produce a number of increasingly complex patterns using three-dimensional blocks to match a two-dimensional design, as quickly as possible. Similar to the standard Block Design subtest, individual items were scored from a minimum of 0 points (failure) to a maximum (which varied across items), based on accurate reproduction and extra points for speed. Item scores were summed and then converted to ability scores using tables in the manual (Elliott, 1990), which takes into account the child's raw score and the difficulty of the items administered. Higher scores are indicative of better performance (i.e., weak CC).

Individuals with autism have been shown to outperform comparison individuals on both tasks, and their superior performance has been attributed to an enhanced ability to presegment the design into its constituent parts, which enables them to locate rapidly either the hidden figure (in the case of the CEFT) or the appropriate blocks from which to reconstruct the pattern (in the case of the Pattern Construction task; see Happé & Frith, 2006, for a review).

General Procedure

Ethical approval for this study was granted by the University of Western Australia's Human Research Ethics Committee, and informed written consent was obtained from parents of all children prior to participation. At each time point, children were seen individually on two occasions, each lasting approximately 1–1.5 hr, either at the family home or at the university. Tests of verbal and nonverbal ability were always administered first followed by tests of specific cognitive skills, the order of which was randomized across participants.

Results

The first section presents preliminary descriptions of data screening and analyses on group matching variables. The subsequent sections present the main results for each cognitive domain separately, including descriptive statistics for the individual tasks and results of analyses assessing developmental persistence and developmental change—two of the primary goals of this study. Where possible, repeated measures analyses of variance (ANOVAs) on individual task scores, with "group" (ASD, typical) as the between-participants factor and "time" as the within-participants factor, were used to examine between-group differences and within-group change. The results of hierarchical regression analyses on the ASD data specifically are also reported to identify potential predictors of change in specific cognitive skills. Aggregate scores for each cognitive domain were used as dependent variables. Also, to minimize the number of predictors in these analyses, only those early developmental variables (age, verbal ability, and nonverbal ability) that were significantly correlated with the dependent variable were entered as potential predictors. The final section reports the results of analyses to determine whether the putative cognitive profile is characteristic of *individual* children with autism.

Preliminary Data Screening

To begin, scores on the set-shifting tasks and the CEFT were reversed so that high scores reflected good performance. Subsequent data screening revealed that the distributions of cognitive variables met assumptions of normality, with the exception of scores on individual ToM tasks, which were significantly positively skewed for children with ASD. Transformations were applied to these ToM variables to try to normalize the data without success. Given the robustness of ANOVA against violations of normality, group differences on individual ToM tasks were analyzed using parametric tests. Also, and not unexpectedly, typical children performed at ceiling on all false-belief tests at Time 2. Analyses concerning developmental changes in ToM skills therefore were restricted to the ASD group. Data screening for outliers (± 3.5 *SD* of the total sample mean) identified one extreme score on the Pattern Construction task at Time 2 (a child with autism). This score was trimmed by replacing it with the value representing 3.5 *SD* above the overall mean (cf. Wilcox, 2002), and subsequent analyses were conducted using the trimmed score (Table 2 displays the untrimmed scores).

Estimates of reliability were ascertained where possible. For ToM, internal consistency across all 14 trials of the false-belief tasks was high for both children with ASD (Time 1: Cronbach's $\alpha = .87$; Time 2: $\alpha = .94$) and typical children (Time 1: $\alpha = .88$). Cronbach's α could not be calculated for typical children's Time 2 ToM scores since performance was at ceiling for this group. The CEFT (response time) showed modest reliability at both time points in the autism (Time 1: $\alpha = .50$; Time 2: $\alpha = .51$) and typical (Time 1: $\alpha = .51$; Time 2: $\alpha = .47$) groups. It was not feasible to calculate reliability estimates for the Pattern Construction task and EF tasks, as most of these measures incorporate stopping rules as part of their administration. High reliability has been reported previously for the Pattern Construction task in preschool and school-aged children (α ranges between .82 and .90; Elliott, 1990), and for set-shifting tasks in autism (Ozonoff, 1995). For each group, there were significant intercorrelations between individual task scores within each cognitive domain at Time 1 (ToM: $r_{\text{mean}} = .66$; EF: $r_{\text{mean}} = .45$; CC: $r_{\text{mean}} = .42$, all $ps < .05$) and at Time 2 (ToM: $r_{\text{mean}} = .55$; EF: $r_{\text{mean}} = .63$; CC: $r_{\text{mean}} = .45$, all $ps < .05$), indicating good convergent validity. Most of these correlations remained significant when the effects of verbal and nonverbal ability were partialled out (all r_{mean}s > .31).

Group Matching

Descriptive statistics for all matching variables are provided in Table 1. There were no significant differences between the ASD and typical groups in terms of chronological age, verbal IQ, or nonverbal IQ at either time point (all $ps > .12$).

Theory of Mind

Table 2 shows mean scores for individual tasks at each time point. As expected, at Time 1, children with ASD obtained significantly lower scores compared with typical children on the first-order unexpected-contents, $F(1, 67) = 47.70$, $p < .001$, $\eta_p^2 = .42$, and unexpected-transfer, $F(1, 67) = 9.84$, $p < .005$, $\eta_p^2 = .13$, tasks, and the second-order unexpected-transfer task, $F(1, 67) = 12.06$, $p < .001$, $\eta_p^2 = .15$. This pattern of group differences was also evident at Time 2: Autistic children performed significantly worse on both first-order false-belief tasks, unexpected-contents: $F(1, 67) = 7.02$, $p < .01$, $\eta_p^2 = .10$; unexpected-transfer: $F(1, 67) = 31.23$, $p < .001$, $\eta_p^2 = .32$, and on the second-order task, $F(1, 67) = 33.33$, $p < .001$, $\eta_p^2 = .34$, than typical children.

Robust aggregate ToM scores for each time point were computed by averaging children's scores from the three individual tasks (see Table 2). A repeated measures ANOVA on aggregate scores showed main effects of group, $F(1, 66) = 42.08$, $p < .001$, $\eta_p^2 = .39$, confirming the results on individual task scores. There was also a significant main effect of time, $F(1, 66) = 108.40$, $p < .001$, $\eta_p^2 = .62$, indicating a significant *improvement* in ToM scores over the 3-year period across groups. The interaction between time and group was not significant, $F < 1$,

Table 2

Mean Scores for Tasks Tapping Each Cognitive Domain in the Autism Spectrum Disorder (ASD; $n = 37$*) and Typically Developing (*$n = 31$*) Groups at Intake and Follow-Up*

	Group			
	ASD ($n = 37$)		Typically developing ($n = 31$)	
Variable	Time 1 *M (SD)* Range	Time 2 *M (SD)* Range	Time 1 *M (SD)* Range	Time 2 *M (SD)* Range
Theory of mind				
First-order unexpected-contents (out of 6)	1.92 (2.13)	4.97 (1.99)	4.84 (1.10)	5.94 (0.36)
	0–6	0–6	3–6	4–6
First-order unexpected-location (out of 6)	1.78 (1.83)	3.32 (2.62)	3.35 (2.30)	5.97 (0.18)
	0–6	0–6	0–6	5–6
Second-order unexpected-location (out of 2)	0.08 (0.28)	1.00 (0.94)	0.61 (0.88)	1.94 (0.25)
	0–1	0–2	0–2	1–2
Aggregate score (maximum score = 4.67)	1.26 (1.25)	3.10 (1.58)	2.94 (1.30)	4.61 (0.17)
	0–4.33	0–4.67	1–4.33	4–4.67
Executive function				
Tower of London (no. trials solved in min. no. moves)	6.33 (2.80)[a]	11.39 (3.21)	10.52 (2.67)[a]	13.94 (2.01)
	2–13	5–18	4–15	9–18
Set-shifting[b] (prop. errors following first sort to criterion)	0.32 (0.07)	0.44 (0.16)	0.23 (0.09)	0.33 (0.08)
	0.18–0.45	0.22–0.80	0.11–0.50	0.20–0.48
Central coherence				
CEFT (score out of 11)	9.73 (1.15)	9.73 (1.02)	8.42 (1.06)	9.42 (1.23)
	7–11	7–11	7–11	6–11
CEFT (s)	5.46 (2.88)	5.79 (2.77)	14.11 (2.56)	8.04 (2.99)
	1.51–12.29	1.21–12.83	8.44–18.41	3.42–15.31
Pattern construction	132.14 (17.64)	136.81 (20.90)	105.64 (12.08)	131.42 (13.54)
	108–182	104–211	75–131	111–173

[a]$n = 36$. [b]Different tasks tapping cognitive flexibility were performed at each time point (see text for details). CEFT = Children's Embedded Figures Test.

although this is most likely due to the ceiling effect present in the typical data. To ensure that the way in which the reported aggregate score was computed did not mask possible deficits on the more difficult second-order ToM tasks, two alternative ToM aggregate scores were constructed. One aggregate score was created by standardizing the scores from each of the three ToM tasks and then deriving the average, and a second score was created by averaging the standard scores of the second-order task and the mean of the two first-order tasks. Subsequent analyses were carried out using these alternative ToM aggregate scores, yet the overall pattern of results remained unchanged.

To examine developmental change within individual children with ASD, the number of children passing and failing each ToM task at Time 1 and Time 2 is presented in Table 3. Success was defined conservatively as ≥ 5 of 6 trials correct (≥ 83%) for each first-order task, and 2 of 2 trials correct (100%) for the second-order task. McNemar's tests showed a significant increase over time in the number of children with ASD passing the first-order

Table 3

Numbers of Children With Autism Spectrum Disorder ($n = 37$*) Passing and Failing Individual Theory-of-Mind Tasks at Intake (Time 1) and Follow-Up (Time 2)*

	Time 2	
Time 1	Fail	Pass
First-order unexpected contents		
Fail: ≤ 4 points	8	23
Pass: 5 or 6 points	0	6
First-order unexpected transfer		
Fail: ≤ 4 points	19	15
Pass: 5 or 6 points	0	3
Second-order unexpected transfer		
Fail: ≤ 1 point	21	16
Pass: 2 points	0	0

unexpected-contents task (16% at Time 1 and 78% at Time 2, $p < .001$), the first-order unexpected-transfer task (8% at Time 1 and 49% at Time 2), and the second-order unexpected-transfer task (0% at Time 1 and 43% at Time 2), all $ps < .001$. All children who passed the first-order ToM tasks at Time 1 also passed these same tasks at Time 2. Of the 9 children who passed either first-order false-belief task at Time 1, all passed the second-order task at Time 2.

A regression analysis using aggregate ToM scores was performed to determine the early predictors of later ToM skills in the ASD group. Time 1 aggregate ToM scores significantly predicted aggregate scores at Time 2 ($\beta = .48$, $R^2 = .23$), $F(1, 36) = 10.57$, $p = .003$. Early verbal ability, $r(35) = .46$, $p < .005$, but not nonverbal ability or chronological age (both $ps > .11$), was significantly associated with later ToM skills. Early receptive-language scores therefore were entered into the model at the second step to determine the additional influence, if any, of this variable on children's later ToM scores. In line with Steele et al. (2003), Time 1 verbal ability explained variance in Time 2 ToM scores over and above Time 1 ToM ($\beta = .30$, $\Delta R^2 = .08$), $\Delta F(1, 34) = 3.88$, $p < .05$, with better vocabulary skills early on predictive of more advanced ToM skills 3 years later. Early ToM scores remained significantly predictive of change in children's ToM scores once variation in early verbal ability had been adjusted for ($\beta = .36$, $p < .05$).

Executive Function

Children's performance on the Tower of London and set-shifting tasks is shown in Table 2. For children's Tower of London scores, a repeated measures ANOVA revealed a significant main effect of group, $F(1, 65) = 37.06$, $p < .001$, $\eta_p^2 = .36$, and time, $F(1, 65) = 125.94$, $p < .001$, $\eta_p^2 = .66$, and a significant Group × Time interaction, $F(1, 65) = 4.69$, $p < .05$, $\eta_p^2 = .07$. Planned comparisons showed that children with ASD obtained significantly lower planning scores than typical children both at Time 1, $F(1, 67) = 40.48$, $p < .001$, $\eta_p^2 = .38$, and at Time 2, $F(1, 67) = 14.55$, $p < .001$, $\eta_p^2 = .18$. Analyses further revealed that children's scores improved significantly over time in both groups, ASD: $t(35) = 10.80$, $p < .001$; typical: $t(30) = 5.64$, $p < .001$. The source of the interaction lay instead with the *extent* of improvements over the 3-year period. Children with ASD made significantly *more* gains in higher order planning (M gain = 5.06, SD = 2.81) than typical children (M = 3.42, SD = 3.37), $F(1, 67) = 4.69$, $p < .05$, $\eta_p^2 = .07$.

How many children showed developmental improvements on the Tower of London task? This question was tackled by determining how many children with ASD showed improvement above and beyond what would be expected given their age and ability. The typical group made an average gain of 3.4 points on the Tower of London task during the 3-year period. The majority of children in the ASD group ($n = 26$; 70% of sample) exceeded this level of improvement (average gain of 5.0 points), compared to 12 typical children (39%).

Separate one-way ANOVAs were carried out on children's set-shifting performance given the use of different tasks at different time points. Children with ASD performed significantly worse than typically developing children on the teddy bear set-shifting task used at Time 1, $F(1, 67) = 18.14$, $p < .001$, $\eta_p^2 = .22$, and on the computerized set-shifting task administered at Time 2, $F(1, 67) = 11.60$, $p < .001$, $\eta_p^2 = .15$, suggesting persistent difficulties with cognitive flexibility.

The significant intercorrelations between Tower of London and set-shifting task scores meant that an EF aggregate score could be computed at each time point by averaging the standard scores for these tasks. Multiple regression analyses demonstrated that individual differences in early EF aggregate scores strongly predicted ASD children's EF performance at follow-up ($\beta = .70$, $R^2 = .49$), $F(1, 35) = 33.22$, $p < .001$. Cross-time correlations revealed that all developmental variables at Time 1 significantly correlated with EF aggregate scores at Time 2, age: $r(35) = .30$, $p < .05$; verbal ability: $r(35) = .48$, $p < .005$; nonverbal ability: $r(35) = .52$, $p < .001$, and therefore all were entered into the model at Step 2. Time 1 EF aggregate scores continued to be significant in this model ($\beta = .66$, $p < .005$), but none of the remaining predictors, however, contributed any unique variance to change in children's EF skills (all $ps > .42$).

Central Coherence

Table 2 shows children's CEFT and Pattern Construction scores. An ANOVA on children's CEFT times revealed significant main effects of group, $F(1, 66) = 124.02$, $p < .001$, $\eta_p^2 = .65$; time, $F(1, 66) = 36.09$, $p < .001$, $\eta_p^2 = .35$; and a significant interaction between group and time, $F(1, 66) = 44.84$, $p < .001$, $\eta_p^2 = .40$. As expected, children with ASD showed a significant advantage on this task, with faster times than typically developing children both at Time 1, $F(1, 67) = 168.29$, $p < .001$, $\eta_p^2 = .72$, and at Time 2, $F(1, 67) = 10.34$, $p < .002$, $\eta_p^2 = .14$. For typical children, the time taken to find the hidden figure reduced significantly over the 3-year period, $t(30) = 7.49$, $p < .001$. Yet this was not the case for children with ASD, whose times remained unchanged, $t < 1$. Comparable results were obtained when median times were used.

An ANOVA on children's Pattern Construction scores revealed main effects of group, $F(1, 66) = 20.75$, $p < .001$, $\eta_p^2 = .24$; time, $F(1, 66) = 60.52$, $p < .001$, $\eta_p^2 = .48$; and a significant Group × Time interaction, $F(1, 66) = 30.87$, $p < .001$, $\eta_p^2 = .32$. Planned comparisons showed that children with ASD performed significantly better than comparison children at Time 1, $F(1, 67) = 50.11$, $p < .001$, $\eta_p^2 = .43$, but failed to maintain their advantage on this task at Time 2, $F(1, 67) = 1.45$, $p = .23$, $\eta_p^2 = .02$. Further within-group analyses showed that, similar to the pattern of results obtained for the CEFT, typical children's scores improved significantly over time, $F(1, 30) = 118.68$, $p < .001$, $\eta^2 = .80$, but ASD children's scores did not, $F(1, 36) = 2.15$, $p = .15$, $\eta^2 = .06$. Examination of individual children's performance over time was addressed in the same way as it had been for children's Tower of London performance. While the ASD group improved, on average, only 4.8 points on the Pattern Construction task between intake and follow-up, typically developing children improved, on average, 25.8 points. Twelve typical children (39%) demonstrated this level of improvement compared with no child in the ASD group.

Central coherence aggregate scores for each time point were created by standardizing the CEFT time and Pattern Construction scores and averaging them. Regression analyses showed that early CC aggregate scores were predictive of change in ASD children's scores over the 3-year period ($\beta = .56$, $R^2 = .31$), $F(1, 35) = 15.85$, $p < .001$. Since Time 2 CC aggregate scores were significantly correlated with initial verbal ability, $r(35) = .64$, $p < .001$, and nonverbal ability, $r(35) = .67$, $p < .001$, but not with age ($p = .10$), the former two Time 1 variables were added to the model at the second step. Nonverbal ability alone contributed an additional 19% of the variance in later CC aggregate scores ($\beta = .50$, $\Delta R^2 = .19$), $\Delta F(1, 34) = 12.66$, $p < .001$, while initial CC scores remained significant ($\beta = .32$, $p < .05$) once variation in verbal and nonverbal ability had been adjusted for.

The fact that the typically developing group improved considerably on both CC measures might imply, paradoxically, that children develop a more pronounced local processing bias over time. Other factors, such as children's maturing nonverbal and executive skills, however, are most likely driving this developmental change. Indeed, performance on both CC measures has been linked previously to concurrent verbal and nonverbal ability (Wechsler, 1999; Witkin et al., 1971), and early nonverbal ability (but not verbal ability) was longitudinally predictive of later performance on the Pattern Construction task, $r(29) = .51$, $p < .005$, and the CEFT, $r(29) = .33$, $p = .06$, in the current sample of typical children. Analyses also revealed a trend for early EF aggregate scores to be associated with later Pattern Construction performance, $r(29) = .32$, $p = .07$, and CEFT performance, $r(29) = .30$, $p = .09$, consistent with the possibility that typically developing children's improving executive skills might enable children to better resist interference from the overall Gestalt and therefore rapidly locate either the hidden figure (in the case of the CEFT) or the blocks necessary to reconstruct the design (in the case of the Pattern Construction task). Since the scores on both CC measures do not represent "pure" indices of local processing, the two CC variables (children's CEFT and Pattern Construction scores) were regressed on nonverbal ability and EF at each time point, and the unstandardized residuals for each variable were saved. For each individual, these residual scores represent the difference between his or her observed score and what would be expected given his or her nonverbal and executive ability. The saved residuals for each variable were then converted to standard scores and averaged to form an aggregate CC score for each time point. These aggregate scores were used in subsequent analyses to address the question of universality of enhanced local processing skills in autism.

Multiple Cognitive Atypicalities at the Individual Level

The results at the group level demonstrate that children with ASD do in fact display the cognitive profile proposed by Happé, Ronald, et al. (2006) and that this profile persists across time. Next, I examined whether *individual* children with ASD showed this profile at both time points.

To examine the universality of cognitive atypicalities, the number of children with ASD that displayed an "atypicality" was calculated for each aspect of cognition using the modified t test developed by Crawford and Howell (1998). This one-tailed test is well suited for comparing single cases to small comparison groups since it treats the statistics of the comparison sample as statistics rather than as population parameters and further provides excellent control over Type I error rates. These authors use Sokal and Rohlf's (1995) formula:

$$t = \frac{X_1 - X_2}{S_2\sqrt{(N_2+1)/N_2}},$$

where X_1 is an individual ASD child's score, X_2 and S_2 are the mean and standard deviation of scores from the typically developing group, and N_2 is the comparison group sample size. Following Crawford and Howell (1998), the *p* value generated by each test is taken to be an estimate of the atypicality of an individual's score. For example, a *p* value of .15 indicates the proportion of the comparison group who would receive a score of similar or greater magnitude than the score obtained for a particular individual. For the purposes of this study, an individual's aggregate score for each of the three domains was classified as "atypical" if the *p* value fell below .15. Note that this definition is functionally similar to previous definitions of an atypicality (Lezak, 1995; see Pellicano, 2007) in which a score is considered to be atypical if it exceeds 1 *SD* above/below the mean score of the typically developing group. Venn diagrams illustrating the results can be found in Figure 1.

At Time 1, 27 of 37 children with ASD (73% of the group) scored below the cutoff ($p < .15$) on the ToM aggregate, 24 children (65%) showed an impairment in EF, and remarkably, all but 1 child (97%) displayed weak CC. How many children showed atypicalities in *all three cognitive domains* at intake (cf. Happé, Ronald, et al., 2006)? More than half of the ASD group (59%) showed co-occurring atypicalities in ToM, EF, and CC. Five children (14%) showed poor ToM and weak CC, 2 children (5%) showed atypicalities in both EF and CC, and the remaining 7 children (19%) displayed weak CC alone (see Figure 1a). One additional child (3%) showed no atypicalities at Time 1.

At Time 2, 26 of 37 children with ASD (70%) fell below the cutoff on the ToM aggregate, 20 children (54%) showed poor EF, and 18 children (49%) showed enhanced local processing. Only 7 children (19%) showed multiple atypicalities in ToM, EF, and CC at Time 2 (see Figure 1b). Of the remaining 30 children, a substantial portion (11 children; 30%) showed joint problems in ToM and EF, 6 children (16%) displayed weak CC combined with poor ToM, 1 child (3%) showed weak CC and poor EF, 2 children (5%) showed ToM difficulties in isolation, 4 children (11%) showed only weak CC, and 1 child (3%) showed EF difficulties alone. Surprisingly, 5 children (13%) showed *no* cognitive atypicalities at follow-up. Of these 5 children, 4 had a clinical diagnosis of autism and 1 child had a diagnosis of PDD–NOS, and 4 of 5 children showed sufficient behavioral symptoms at follow-up to meet the cutoff for autism on the ADOS–G (Lord et al., 1999). Among the 22 children who showed all three atypicalities at Time 1, 5 children continued to display this profile at Time 2, 11 children showed persistent difficulties in ToM and EF, 2 children showed atypicalities in ToM and CC, 3 children showed ToM

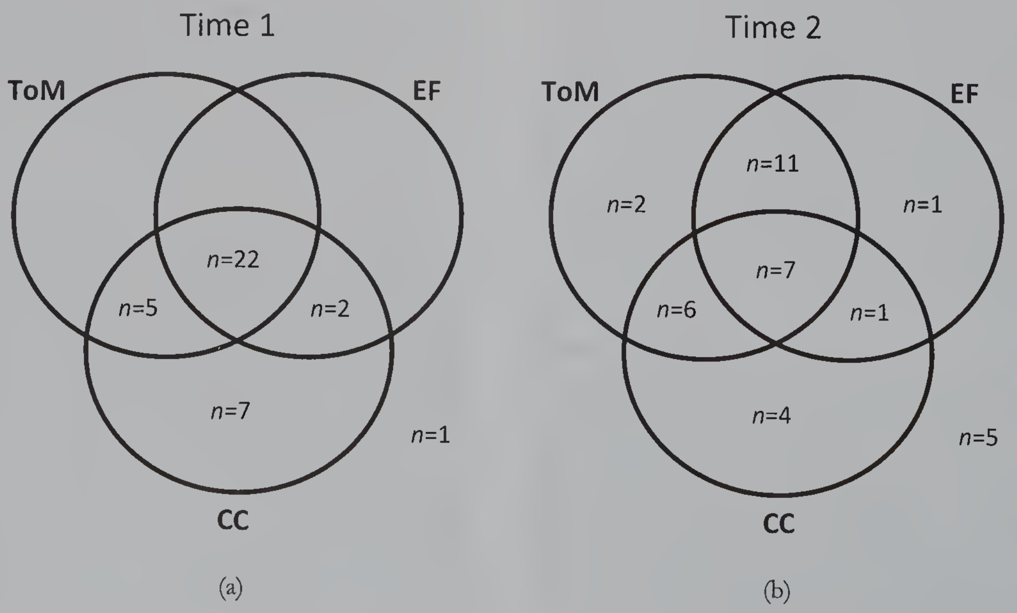


Figure 1. The overlapping circles in these Venn diagrams represent the potential combinations of atypicalities in ToM, EF, and CC. The region in the center represents co-occurring atypicalities in all three cognitive domains: the putative cognitive phenotype of autism (cf. Happé, Ronald, & Plomin, 2006). The numbers inside each region represent the number of children with an autism spectrum disorder condition showing that particular atypicality alone or combination of atypicalities at (a) Time 1 and (b) Time 2.
Note. ToM = theory of mind; EF = executive function; CC = central coherence.

problems in isolation, and 1 child showed no atypicalities.

It is possible that, among those children who failed to demonstrate atypicalities in one or more cognitive domains, such atypicalities were in fact present in some or all of these children, albeit to a lesser degree. To investigate this further, additional analyses were carried out using a less conservative definition of atypicality, where an individual's aggregate score for each cognitive domain was identified as ''atypical'' if the p value fell below .50. (Note that this is a lenient definition of atypicality since the p value estimates the point at which 50% of the typical population would obtain scores lower than the score obtained for a particular individual; cf. Crawford & Howell, 1998.) According to the main universality analyses using the conservative criterion ($p < .15$; see Figure 1), at Time 1, 10 children showed no problems in ToM, 13 children demonstrated no EF difficulties, and 1 child failed to show weak CC. Subsequent analyses using the more lenient criterion ($p < .50$) revealed that all of these cases showed atypicalities in these domains (ToM: all cases $\leq p = .48$; EF: all cases $\leq p = .43$; CC: all cases $\leq p = .43$). At Time 2, the central, more conservative analyses revealed that 11 children showed no ToM difficulties, 17 children showed no EF problems, and 19 children failed to show weak CC. Again, when the more lenient criterion was applied, all children were classified as atypical on each aspect of cognition (ToM: all cases $\leq p = .38$; EF: all cases $\leq p = .46$; CC: all cases $\leq p = .45$), suggesting that the entire sample of children with ASD appeared to show the putative cognitive profile, albeit in subtle form.

Discussion

Developmental Persistence

Relative to age- and ability-matched typically developing children, children with ASD showed difficulties in false-belief understanding, problems with higher order planning and cognitive flexibility together with capabilities in processing local information *at both time points*. This finding supports Happé, Ronald, et al.'s (2006) claim that there is a specific profile of coexisting cognitive atypicalities in autism, which shows continuity with development. Yet testing the claims of Happé, Ronald, et al.'s multiple-deficits model required not only analysis of the group as a whole but also examination of the performance of individual children with ASD. Such analysis demonstrated that the putative profile was not universal at either time point and became markedly less pervasive at follow-up. At intake, weaknesses in false-belief attribution and aspects of EF were found to be less pervasive than atypicalities in CC, resulting in just over half of the ASD group displaying coexisting atypicalities in all three domains. At follow-up, a dramatically different pattern emerged: None of the atypicalities were entirely pervasive, and only a minority of children displayed the complete profile. Inspection of the group means at Time 2 revealed that the distributions overlapped greatly, but particularly so for the CC tasks. Indeed, group differences on the CEFT were attenuated at Time 2, and had disappeared altogether on the Pattern Construction task.

Measurement issues might have contributed to the failure to demonstrate the universality of this cognitive profile in two ways. First, the Time 2 analyses on ToM were complicated by the presence of ceiling effects in the typically developing group, which could have influenced the magnitude of the group difference and, in turn, the number of individual children with ASD showing ToM difficulties. Even in the event that this was true, it is difficult to see how this explanation alone could explain the minority of children with ASD ($n = 7$) showing all atypicalities at follow-up since this ceiling effect should not have affected the results concerning atypicalities in EF and CC, which together accounted for less than half of the group.

Second, the cognitive measures were matched neither for discriminative power nor for reliability (cf. Chapman & Chapman, 1973), rendering it possible that the tasks were not sufficiently sensitive to detect subtle cognitive atypicalities. Internal consistency was moderate to high for cognitive tasks, which remained stable across time points, and there was good convergent validity for each cognitive domain at intake and follow-up. Reliability was not evaluated for executive measures, which has been reported to be low in typical samples (e.g., Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001) and high in autism samples (e.g., Ozonoff, 1995). The results of these analyses therefore should be treated with some caution.

Despite these limitations, the fact that the majority of children failed to show multiple cognitive atypicalities at follow-up combined with the presence of some children who showed *no* cognitive atypicalities is problematic for Happé, Ronald, et al.'s (2006) candidate multiple-deficits account, which attempts to provide an integrated causal explanation for the pathogenesis of autism. It is noteworthy that analyses using a more lenient definition of atypicality demonstrated that *all* children with ASD showed the putative cognitive profile, which is encouraging for Happé, Ronald, et al.'s account. It is worth emphasizing, however, that according to this lenient definition of atypicality ($p < .50$), one should expect to find 50% of individuals in the population from which the comparison sample was drawn also to receive a score of similar or greater magnitude than that observed for individuals with ASD (Crawford & Howell, 1998). In such instances, one would be cautious in classifying these typical children as showing an atypicality, and therefore one must be equally cautious of doing so in the case of ASD.

Certainly, the findings from these supplementary analyses serve to highlight the challenges in defining atypicality, especially when examining individuals with developmental conditions, whose cognitive skills are unlikely to be "all or none." Rather than "atypicality" being conceptualized categorically, it might be beneficial to conceive of the three cognitive domains as dimensions located (orthogonally) within a multivariate space (Happé & Ronald, 2008; Happé, Ronald, et al., 2006). The extent to which a person with autism shows a particular cognitive atypicality therefore should vary according to the place he or she occupies on that dimension, which in turn, should relate directly to the degree and nature of the behavioral symptoms it purports to explain. This elaboration is appealing as it could potentially account for the phenotypic heterogeneity in autism. It would be of interest, therefore, to demonstrate empirically that individual differences in cognitive skills are indeed significantly related to individual differences in specific aspects of the behavioral phenotype—both concurrently and longitudinally. Despite its appeal, it nevertheless remains a challenge for this account to explain the absence of the suggested cognitive atypicalities in 4 children with ASD who have sufficient behavioral symptoms at follow-up. It is plausible that the emergence of cognitive atypicalities at some point in development could be sufficient to cause behaviors severe enough to persist beyond the cognitive atypicalities themselves.

Developmental Progress in Autistic Children's Cognitive Skills

An understanding of such putative processes necessarily demands a fuller developmental perspective than is currently offered. The second main finding from this study was that were considerable developmental improvements in autistic children's ToM and EF, but no changes in local processing. Children with ASD made substantial progress in their false-belief understanding over the 3-year period to the extent that more than one third of the group succeeded on advanced false-belief tasks at follow-up. These findings present a more optimistic picture of autistic children's developing ToM than earlier longitudinal studies, which reported negligible progress in children's false-belief understanding (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994). Notably, children in these early investigations were older and less able than participants sampled both in the current study and in two recent longitudinal studies, which also demonstrated improvements in some aspects of children's ToM skills (Serra et al., 2002; Steele et al., 2003), highlighting the possibility that improvements, should they occur, might be most apparent early on during development (cf. Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). The greater emphasis in recent years on early behavioral intervention for young children with autism might also account for the greater developmental progress reported in more recent studies compared to earlier ones.

Children's planning capacity also improved significantly over the 3-year period—at a strikingly faster rate than that of typically developing children. These findings go against the two existing prospective studies on EF in autism, which reported progress in EF neither in very young children with autism (Griffith et al., 1999) nor in adolescents with autism (Ozonoff & McEvoy, 1994). Again, differences in sample selection could explain the opposing pattern of results since both of these studies included individuals with autism less able than the sample of autistic children assessed here. Interestingly, and akin to the current findings, Happé, Booth, Charlton, and Hughes (2006) found that the EF performance of older participants with autism (*M* age = 13 years 2 months) surpassed that of younger children (*M* age = 9 years 2 months), suggestive of age-related gains in EF. Furthermore, although the authors did not find evidence of specific problems in planning ability, they nonetheless showed that this EF component showed the greatest age-related improvements in the autism group. The current longitudinal data strengthen Happé, Booth, et al.'s cross-sectional findings, and indicate that EF problems become *less marked with age*, at least for cognitively able children with autism.

Several factors might explain this boost in ASD children's ToM and executive skills over the 3-year period. The first of these is language. Regression analyses showed that verbal ability contributed significant variance in later ToM scores over and above variance already accounted for by early ToM performance. The important influence of language during children's developing ToM is consistent with an abundance of cross-sectional (e.g., Fisher, Happé, & Dunn, 2005; Happé, 1995) and some longitudinal (Steele et al., 2003; Tager-Flusberg & Joseph, 2005) work, showing that language ability, and especially grammatical ability, is significantly related to autistic children's performance on ToM tasks. Alongside these findings, the current data provide firm evidence that early language skills might play an important role in facilitating autistic children's understanding of the representational nature of mind.

Early verbal ability was, however, unrelated to children's developmental improvements in planning ability. This finding was unexpected since there are both empirical and theoretical reasons to anticipate a link between language and children's developing executive skills. Russell (1996) argued that (internal) language ability plays a key role in executive control over action. Specifically, success on executive tasks, such as the Tower of London task, should be bolstered by the use of inner speech, that is, if the child forms a *verbal representation* of the sequence of moves necessary to solve the problem in the minimum number of moves. Two studies have shown experimentally that reduced executive control in autism might indeed result from an inherent failure to use language to regulate and control one's actions (Joseph, Steele, Meyer, & Tager-Flusberg, 2005; Whitehouse, Maybery, & Durkin, 2006). If language is integral to success on EF tasks, then it is puzzling that early language skills were unrelated to developmental *improvements* on such tasks in the current study. Perhaps children's use of compensatory strategies, which might be nonverbal in nature, becomes increasingly practiced over time, affording gradual gains in executive control.

The second factor that could explain the enrichment in autistic children's developing ToM and EF is social contact (Hughes, 1998b). Work with typically developing children shows that early peer interactions, particularly those involving pretense, foster the development of ToM (Hughes & Dunn, 1998; see Dunn, 2004, for a review). Likewise, social relations are likely to influence children's developing executive skills since negotiating everyday social interactions requires children to regulate their own behaviors (Luria, 1966; see also Hughes, 1998b). The majority of autistic children in the current study attended mainstream classrooms ($n = 34$), and therefore encountered frequent (and varying) models of age-appropriate social interactions. Evidence suggests that cognitively able children with autism seek out their nonautistic peers (Sigman & Ruskin, 1999). Engaging with more advanced social partners on an everyday basis therefore might have bootstrapped the development of these children's mental-state reasoning and executive control (Vygotsky, 1978). Indeed, increased social contact during this period of development might also account for the *enhanced* rate of acquisition of EF skills compared to typical children. EF shows a protracted development trajectory, which renders it especially susceptible to environmental influences. Recent findings suggest that EF skills are indeed malleable in young children (Diamond, Barnett, Thomas, & Munro, 2007). It therefore remains possible that attendance in mainstream schools provided an enriched social environment with ample opportunities for autistic children to "exercise" their EF skills, therefore boosting the rate of acquisition of such skills. Identifying which aspects of social relations, if any, partially mediate the development of children's EF and ToM skills is a worthy line of investigation since this should have implications for the remediation of atypicalities in these cognitive domains.

A third reason for the improvements in EF and ToM is that the development of these functions is in fact inextricably linked. There is strong evidence that ToM and EF are closely tied in typically developing children (e.g., Carlson & Moses, 2001; Hughes, 1998a, 1998b) and in autism (e.g., Pellicano, 2007). The nature of this link, however, is controversial. Some authors have proposed that executive skills are a prerequisite for the later development of ToM (Moses, 2001; Russell, 1996), while others claim that a representational understanding of mind is necessary for the later control of goal-directed, purposeful action (Perner & Lang, 1999). Further still, the parallel progression of ToM and EF might be due to a third factor: that both skills are mediated by adjacent structures in the prefrontal cortex (Ozonoff et al., 1991). Research on the overlap between these two domains in autism is scarce yet analysis of the longitudinal links between EF and ToM in the current sample of children with ASD (see Pellicano, 2010) points toward a developmental relationship in one direction only. Earlier EF skills were found to be longitudinally predictive of change in autistic children's ToM test performance, independent of age, verbal ability, nonverbal ability, and early ToM skills but there were no significant predictive relations in the opposite direction. Akin to work with typically developing children, these findings find favor with Russell's account that EF plays an important role in the advancement of autistic children's ToM skills (see also Pellicano, 2007). Nevertheless, the possibility of a functional link between EF and ToM is potentially inconsistent with Happé, Ronald, et al.'s (2006) model, which considers the co-occurring cognitive atypicalities in autism to be independent and genetically distinct, and is therefore unlikely to support any notion of one atypicality emerging from another.

Despite significant improvements in ToM and executive skills, children experienced no changes in local processing over the 3-year period. Children with ASD were no quicker to find the hidden figure on the CEFT than they were 3 years earlier, and their scores on the Pattern Construction task failed to change over time. This lack of change is in striking contrast to the significant gains made by typically developing children evidenced on both weak CC tasks. Could the lack of an improvement be explained by a ceiling effect in the ASD group? This explanation might account for children's CEFT performance, but it is unlikely that a ceiling effect could account for children's equally limited improvement on the Pattern Construction task since the majority of children's scores were well below the maximum score.

One alternative explanation is that the development of local processing takes strikingly different courses in autism and typical development. This suggestion is consistent with one competing account of perceptual atypicalities in autism, enhanced perceptual functioning (Mottron & Burack, 2001), which proposed that such peaks derive from the overdevelopment of basic perceptual processes, including local processing. Rather than being present from birth, superior local processing emerges early on during the course of development in response to diminished processing of higher order operations. In line with Mottron and Burack's (2001) claim, the current findings suggest that these skills are early emerging and initially accelerated in autism. Over time, it appears that the trajectories of children with ASD and typical children converge as typical children's performance on tasks tapping local processing "catches up" to that of autistic children although potentially via a different route (see Results for details).

The current data also highlight the possibility that the developmental trajectory of local processing in autism might be qualitatively distinct from the trajectories of other, higher cognitive domains (ToM and EF), which appear to mature later and progress somewhat in parallel. Do these findings fit with Happé, Ronald, et al.'s (2006) theoretical position? In its current form, the model says little about the ways in which these cognitive atypicalities unfold with development and which sort of factors influences their progression. The possibility of distinct developmental trajectories for some aspects of cognition might be taken as evidence to support a fractionable cognitive characteristics account (Happé & Ronald, 2008). Although Happé, Ronald, et al. hinted at the possibility of potential "interactions" between cognitive domains, the nature of any such interactions remains unspecified. Further analysis of the current longitudinal data (see Pellicano, 2010) suggest that interactions do in fact exist, specifically in which early domain-general skills (EF and CC) play a crucial role in shaping the developmental trajectory of autistic children's emerging ToM. Certainly, further elaboration of Happé, Ronald, et al.'s model is required to explain the continuities and discontinuities in the development of core cognitive skills, the specific *developmental* relationships between cognitive domains, and the potential factors (e.g., verbal ability, social interaction; Hughes, 1998b) that might mediate or moderate such development.

Conclusion

This is the first prospective study to investigate the development of *multiple* cognitive atypicalities in ASD. Several tasks tapping ToM, EF, and CC were administered to the same samples of children at two different time points separated by 3 years. The findings showed that cognitively able children with ASD, as a group, show coexisting atypicalities in important aspects of ToM, EF, and CC, relative to typically developing comparison children, and that this profile persists over time—precisely in the way Happé, Ronald, et al. (2006) proposed. Their multiple-deficits model goes far beyond previous single-deficit models by invoking several core underlying atypicalities that, together, underpin the symptoms of autism—arguably, a much more realistic position given the heterogeneity present in the condition. Yet not all children showed this particular cognitive profile. Nor did children's cognitive skills remain static and unchanging. In fact, they experienced considerable gains in those aspects of cognition that typically present them with the most difficulty—progress that is encouraging and perhaps unsurprising in developing cognitive systems, but is nonetheless difficult to predict from Happé, Ronald, et al.'s model in its current form. Whether this pattern of findings will be similar for children with autism who have additional learning difficulties remains a crucial question for future research. The current findings nevertheless stress the need for theorists to take seriously the *developing* cognitive phenotype of autism.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition, 21*, 37–46.

Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task. *Journal of Child Psychology and Psychiatry, 42*, 551–556.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., et al. (1994). A case–control family history study of autism. *Journal of Child Psychology & Psychiatry, 35*, 877–900.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79*, 380–385.

Comerford, B. (2005). *Can inner speech deficits explain executive function deficits in autism?* Unpublished thesis, University of Western Australia.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist, 12*, 482–486.

de Villiers, J. (2000). Language and theory of mind: What are the developmental relationships. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 83–123). Oxford, UK: Oxford University Press.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Dunn, J. (2004). *Children's friendships: The beginning of intimacy*. Oxford, UK: Blackwell.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Elliott, C. D. (1990). *Differential ability scales*. New York: Psychological Corporation.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 46*, 409–419.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Blackwell.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development, 70*, 817–832.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*, 843–855.

Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: Examining profiles across domains and ages. *Brain and Cognition, 61*, 25–39.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*, 5–25.

Happé, F., & Ronald, A. (2008). The "fractionable autism triad": A review of evidence from behavioural, genetic, cognitive, and neural research. *Neuropsychology Review, 18*, 287–304.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation of autism. *Nature Neuroscience, 9*, 1218–1220.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources.

Hill, E. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*, 26–32.

Holroyd, S., & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders, 23*, 379–386.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*, 1326–1339.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*, 1026–1037.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental Psychology, 29*, 498–510.

Joseph, R. M., Steele, S. D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory. *Neuropsychologia, 43*, 1400–1411.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., et al. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42*, 261–270.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule–Generic*. Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview–Revised. *Journal of Autism and Developmental Disorders*, *24*, 659–685.

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Mandy, W. P., & Skuse, D. H. (2008). Research review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *Journal of Child Psychology and Psychiatry*, *49*, 795–808.

Moses, L. J. (2001). Executive accounts of theory of mind development. *Child Development*, *3*, 688–690.

Mottron, L. & Burack, J. (2001). Enhanced perceptual functioning in the development of persons with autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131–148). Mahwah, NJ: Erlbaum.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology*, *9*, 491–500.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology*, *6*, 415–431.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, *32*, 1081–1105.

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*, 974–990.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology*, *46*, 530–544.

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: ''Weak'' central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology*, *18*, 77–98.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 51–87.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*, 337–344.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–137.

Perner, J., & Wimmer, H. (1985). ''John *thinks* that Mary *thinks* that . . . '': Attributions of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*, 224–260.

Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale–Revised*. Wood Dale, IL: Stoelting.

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 691–699.

Russell, J. (1996). *Agency: Its role in mental development*. Hove, UK: Erlbaum.

Rutter, M. (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry*, *24*, 513–531.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Serra, M., Loth, F. L., van Geert, P. L. C., Hurkens, E., & Minderaa, R. B. (2002). Theory of mind in children with ''lesser variants'' of autism: A longitudinal study. *Journal of Child Psychology and Psychiatry*, *43*, 885–900.

Shallice, T. (1982). Specific impairments in planning. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *298*, 199–209.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development*, *64* (Serial No. 256).

Sokal, R. R., & Rohlf, J. F. (1995). *Biometry*. San Francisco: W.H. Freeman.

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Developmental change in theory of mind abilities in children with autism. *Journal of Autism and Developmental Disorders*, *33*, 461–467.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). Oxford, UK: Oxford University Press.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. New York: Psychological Corporation.

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry*, *47*, 857–865.

Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA. *Journal of Clinical Child and Adolescent Psychology*, *31*, 399–412.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. (1971). *A manual for the Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities in individuals with autism, individuals with mental retardation and normally developing individuals. *Psychological Bulletin*, *124*, 283–307.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1417–1430

Gender Differences in Spatial Ability of Young Children: The Effects of Training and Processing Strategies

David Tzuriel and Gila Egozi
Bar Ilan University

A sample of 116 children (*M* = 6 years 7 months) in Grade 1 was randomly assigned to experimental (*n* = 60) and control (*n* = 56) groups, with equal numbers of boys and girls in each group. The experimental group received a program aimed at improving representation and transformation of visuospatial information, whereas the control group received a substitute program. All children were administered mental rotation tests before and after an intervention program and a Global–Local Processing Strategies test before the intervention. The results revealed that initial gender differences in spatial ability disappeared following treatment in the experimental but not in the control group. Gender differences were moderated by strategies used to process visuospatial information. Intervention and processing strategies were essential in reducing gender differences in spatial abilities.

Spatial ability has a vital role in our daily interaction with environment, such as navigation, recognizing and manipulating objects, academic tasks, and recalling locations. Spatial ability is one of the several relatively autonomous human intellectual competencies (Gardner, 1983) and is considered essential to mathematics and scientific thinking (Delgado & Prieto, 2004), performance on standardized tests (e.g., the SAT; Casey, Nuttall, Pezaris, & Benbow, 1995), representing and manipulating of information, and problem solving. Proficiency in visuospatial ability has long been associated with success in cognitively demanding educational tracks and occupations such as engineering, architecture, physics, chemistry, and surgery (e.g., Snow & Yalow, 1982; Sorby & Baartmans, 2000) and is a salient characteristic of physical scientists (Gohm, Humphreys, & Yao, 1998; Humphreys, Lubinski, & Yao, 1993). Visuospatial ability is not routinely taught by schools and thus is not often developed and assessed in ways that influence students' educational and career plans (Webb, Lubinski, & Benbow, 2007). Adding spatial ability to identification procedures of talented students (currently restricted to mathematical and verbal ability) could uncover a neglected pool of math–science talent and holds promise for refining our understanding of intellectually talented youth.

Gender differences in spatial ability are well documented in the scientific literature (e.g., Coluccia & Louse, 2004; Halpern, 2007; Kimura, 1999; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005; Voyer, Voyer, & Bryden, 1995). Early researchers in this area have traditionally reported a male advantage over female on standard tests of spatial ability, at least after adolescence (Maccoby & Jacklin, 1974). Recent research, however, showed that gender differences emerge around the time children enter kindergarten or begin first grade, which may be as early as children can reliably perform tasks that assess visuospatial abilities. Levine, Huttenlocher, Taylor, and Langrock (1999) found that, on average, preschool boys are more accurate than girls at spatial tasks that measure that accuracy of spatial transformations ($d = .31$) and score higher on the Mazes subtest of the Wechsler Preschool and Primary Scale of Intelligence ($d = .30$). They concluded that gender differences in favor of boys are present on spatial tasks by age 4½.

Meta-analytic studies suggested that spatial ability is not a unitary construct, and that the magnitude of gender differences deeply depend on a task's demands (Coluccia & Louse, 2004; Kimura, 1999; Linn & Petersen, 1985; Voyer et al., 1995). However, those reviews provided substantial evidence that the largest and the most robust gender

---

Correspondence concerning this article should be addressed to David Tzuriel, School of Education, Bar Ilan University, Ramat Gan 52900, Israel. Electronic mail may be sent to tzuried@mail.biu.ac.il.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0008

differences arise on spatial tasks involving mental rotation, like the classical mental rotation test developed by Shepard and Metzler (1971). Gender differences in mental rotation tests are typically associated with some of the earliest disparities in strategies used by boys and girls in processing visual–spatial information (Cooper, 1976; Kramer, Leopard, Ellenberg, & Share, 1996). Recent studies with infants, using habituation techniques, demonstrated that already by the age of 3–5 months, there were gender differences in mental rotation (Moore & Johnson, 2008; Quinn & Liben, 2008). Quinn and Liben (2008) provided evidence showing that male infants were more likely than female infants to recognize as equivalent a two-dimensional figure rotated in the same way as was the figure shown during a familiarization phase. The male infants were also more likely than female infants to perceive a mirror image of that figure as novel. Similarly, Moore and Johnson (2008) demonstrated gender differences with 5-month-old infants using two-dimensional object representing three-dimensional object. The male infants discriminated the habituation objects from its mirror-image test objects whereas female infants treated both objects similarly. In both studies, the investigators conclude that in spite of the fact that gender differences in mental rotation are already apparent at infancy, more research should be carried out to explore the interacting factors that influence the development of mental rotation.

Several researchers reported that global–local strategies were inherently related to mental rotation performance (Bethell-Fox & Shepard, 1988; Bryden, George, & Inch, 1990; Johnson, 1990). Small effect sizes were reported for simple commonly used tasks whereas in more difficult tasks that require generating an image and maintaining it in working memory, results vary depending on the complexity of the image to be generated and the specific nature of the task, with observed differences ranging between *d* = .63 and *d* = .77 (Loring-Meier & Halpern, 1999). Mental rotation tasks that require maintaining a three-dimensional figure in working memory while simultaneously transforming it show very large gender differences, somewhere between 0.90 to 1.0 *SD* (Masters & Sanders, 1993; Nordvik & Amponsah, 1998), although some researchers have reported smaller effect sizes (Voyer et al., 1995).

Males' outperformance of females on measures of visuospatial abilities has been implicated as contributing to gender differences on standardized exams in mathematics and science (Halpern et al., 2007). An evolutionary account of gender differences in mathematics and science (e.g., Geary, 2007; Gur & Gur, 2007; Haier, 2007) supports the conclusion that although gender differences in math and science performance have not directly evolved, they could be indirectly related to differences in interests and specific brain and cognitive systems. Newcombe (2007) challenged the evolutionary approach, first, by showing logical inconsistencies and gaps in biological sex differences research and, second, by demonstrating that these differences are not immutable and can be eliminated by proper intervention.

In the present study, we address four important interrelated questions: (a) Do young children in first grade show gender differences on mental rotation tasks at an appropriate level of difficulty for their age? (b) To what degree does training of representation and transformation of perceptual stimuli, by viewing them from different perspectives, reduce initial gender differences? (c) Is the magnitude of the hypothesized gender differences influenced or moderated by global versus local strategies of representing and transforming visuospatial information? (d) What are the interactive effects of gender, training, and task difficulty (degree of rotation and test level) on performance of mental rotation tasks?

For the purpose of this study, we constructed a mental rotation test, which is based on the *Mental Rotation* subtest of the *Cognitive Modifiability Battery* (CMB; Tzuriel, 1995, 2000a, 2000b; Tzuriel & Egozi, 2006) and an intervention program aimed at training global strategies of representing and transforming of spatial information based on Wheatley's (1996) ''Quick Draws'' activities. According to Piaget and Inhelder (1956), these abilities are vital components in mental rotation performance. Representing and transforming spatial information requires to some degree an extension of working memory capacity proved to be essential for mental rotation performance. Coluccia and Louse (2004) in their extensive meta-analytic review have reached a conclusion that gender differences in spatial orientation emerge only when tasks require a high load of visuospatial working memory. Thus, males would show better performance in spatial orientation because of their larger visuospatial working memory span. When the orientation task does not involve a high load in visuospatial working memory, gender differences would disappear. Similarly, Feng, Spence, and Pratt (2007) showed that improvement in performance on a mental rotation task, following training composed of playing an action video game, was associated with improvement in spatial attention. Furthermore, following training, females showed higher improvement than males on both spatial attention and mental rotation tasks. Vasta, Knott, and Gaze (1996) showed also that gender differences were eliminated by carefully conceptualized training.

In congruence with other researchers (Feng et al., 2007; Newcombe, Mathason, & Terlecki, 2002), our claim is that it is more important to determine whether gender differences may be reduced or eliminated by training than merely documenting them. Baenninger and Newcombe (1989) reported in their meta-analytic study that spatial ability can indeed be improved with training, with effect sizes ranging from $d = .40$ to $d = .80$, depending on the length and specificity of the training. However, no differential improvements were reported for boys and girls. A more recent meta-analysis of the effects of training and experience on spatial skills confirms the earlier findings (Marulis, Lui, Warren, Uttal, & Newcombe, 2007). However, the effects of training were similar for males and females; that is, both groups benefited about equally from the training; no evidence was found to show that the gap was closed or widened by training.

In a recent study by Tzuriel and Egozi (2006), gender differences in mental rotation were investigated among 5- to 6-year-old kindergarten children. The objectives were to investigate the interactive effects of gender, treatment, and task characteristics on spatial abilities tasks using a *dynamic assessment* (DA) procedure (Haywood & Lidz, 2007; Lidz & Elliott, 2000; Sternberg & Grigorenko, 2002; Tzuriel, 2001). The treatment was composed of teaching children to preserve the whole pattern of a model at different positions in space. This teaching strategy focuses on a global-holistic approach based on pattern conservation of the model and rotating it in the space as a whole unit, without fragmenting its components. Clinical experience showed the global-holistic strategies to be more efficient than analytic strategies, as children could use global-holistic strategies to circumvent the difficulty of defining verbally the problem's components. A verbal analytic strategy was given as a secondary strategy to support children who might have difficulty with the holistic approach. The findings revealed a significant improvement in mental rotation in both gender groups. A differential treatment effect on boys and girls was evident, however, when task characteristics (e.g., degree of rotation, pattern complexity) were involved. Contrary to the hypotheses, *boys responded more effectively to intervention at difficult levels of the tasks and improved their mental rotation more than did girls.* The differential improvement of boys and girls was attributed to different strategies used by them (e.g., global), as evidenced by clinical observations during the teaching phase. In spite of the fact that all children were offered similar strategies, boys tended to use a global approach based on *conservation of patterns and rotation of the unit as a whole,* whereas girls showed a detailed analytic approach based on verbalization of location.

These findings and the clinical evidence showing the effectiveness of a global-holistic perceptual approach in representing visuospatial information were the basis of the current study. Unlike the short 1-hr intervention (teaching phase of a DA procedure) used earlier (Tzuriel & Egozi, 2006), in the current study we used a more intensive intervention approach composed of eight small group sessions (45 min each).

Following Tzuriel and Egozi's (2006) study, our main hypothesis was that initial gender differences in mental rotation will be reduced by the use of adequate training based on developing cognitive processing strategies of representation and transformation of visuospatial information. We also hypothesized that boys will show higher global strategies in processing visuospatial tasks than girls who will tend to show local strategies, and that the initial gender differences will disappear when Global–Local Processing Strategies (GLPS) are considered. With regard to the intervention program, we hypothesized that it would be more effective for high-level than low-level tasks (i.e., as degree of rotation increases).

Method

Participants

The sample was composed of 116 first-grade children (58 girls and 58 boys), recruited from Israeli public schools. Most children were of Jewish ethnic origin (89.66%) and the rest were Israeli-Arabs (10.34%). Children ranged in age from 6 years to 7 years 2 months (M = 6 years 7 months, SD = 3.71 months), with no significant age differences between boys and girls, $F(1, 113) = 1.81$, *ns*. All children were right-handed, with no history of developmental, neurological, or learning problems. The reason for including only right-handed children is based on recent neurological research showing different hemispheric activity among left- and right-handed individuals (Kim et al., 1993). Some researchers suggest controlling this variable when studying gender differences in spatial ability (e.g., Kramer et al., 1996). The sampling process was carried out in three stages: (a) random selection of 10 of 34 public primary schools whose principals agreed to apply for an intervention program for the development of spatial abilities, (b) random selection of one first-grade class in each school, and (c) a random selection of 6 boys and 6 girls in each class. The parents were required to consent to the inclusion of their child in the study and provide basic information on the child's developmental history. During data gathering, 4 of the original sample of 120 children could no longer participate in the study because their families moved from the area.

A sample of 3 boys and 3 girls from each class were randomly assigned to experimental and control groups. The experimental group consisted of 60 children (30 each, boys and girls) and the control group consisted of 56 (28 each, boys and girls). Both groups were similar on socioeconomic status (SES) indices of parent occupation, $\chi^2(4) = 3.10$, *ns*, and years of education, $\chi^2(4) = 3.68$, *ns*. In both groups, about 50% of the parents graduated high school and held a professional occupation. In view of Levine et al.'s (2005) findings that gender differences are sensitive to SES differences, the similarity of the experimental to the control group in the current sample seems to be satisfactory. No significant differences were found between the groups on a composite score of the WISC–R95 Vocabulary and Concept Formation subscales, $F(2, 112) = .58$, *ns*.

Measures

Spatial Relations (SR) subtest from the Primary Mental Abilities–Children's Version (Thurstone & Thurstone, 1965). The SR is a standard measurement for the ability to visualize rotated objects in two-dimensional space and to identify the relation among them. The SR is composed of 27 items and four examples. In each item, the child is presented with a model figure that is a part of a square and has to determine which of the four given alternatives completes the square. The alternatives are positioned so that children must mentally rotate the parts to complete the square. A score of 1 is given for each correctly solved item.

Although standardized administration of the SR is in groups, with a time limit of 4 min, in the present study it was administered individually and without a time limit, to avoid test stress and anxiety. Johnson and Meade (1987) reported a split-half reliability of .89, using the Spearman–Brown's formula among children in Grades 1–4. Correlations with other spatial ability tests ranged from .44 to .63.

The Windows Test (WT). Construction of the WT was based on the *Mental Rotation* subtest of the *CMB* (Tzuriel, 1995, 2000a, 2000b; Tzuriel & Egozi, 2006). Designed for children in first to fourth grades, it includes three difficulty levels (WT1, WT2, and WT3). In order to examine the adequacy of the WT for young children, a pilot study was carried out on Grade 1 children ($n = 32$). Based on the preliminary results where a floor effect was observed on the WT3 level, it was decided to administer only the first two levels (WT1 and WT2). For each test level, there are parallel versions that are administered in the pre- and postintervention phases, respectively.

In both test levels, children are presented with model figures of "houses with windows" arranged in a 3 × 3 patterns (nine windows); some windows are closed (blackened) whereas others are open. On turned-about houses, taking into account their rotation in space (see Figure 1), children are asked to mark the identical closed windows. The main difference between WT1 and WT2 is the clue for rotation, which is much more concrete and articulated in WT1 (red roof) than in WT2 (red basis line). For each level, there are 18 problems arranged according to the *degree of rotation* needed to solve the tasks (45°, 90°, and 180°), *complexity* (2, 3, and 4 closed windows), and *symmetry* (half are symmetrical and half are asymmetrical; see Figure 1). The unique design of the WT allows separate analyses of the child's responses according to test level, degree of rotation, complexity, and symmetry of the patterns. In the current study and for the sake of simplicity, we analyzed only the test level and degree of rotation.

Each correctly solved window is given a score of 1 (maximal score = 54). Cronbach alpha reliabilities in the pilot study were .79 and .82 for the WT1 and WT2, respectively. Test equivalency reliabilities for the WT1 and WT2 were .82 and .87, respectively. The Cronbach alpha reliabilities on the current Grade 1 sample, for pre- and postintervention, were: WT1 = .78 and .73, respectively; WT2 = .76 and .74, respectively. Test–retest correlations between pre- and postintervention in the experimental group were .60 and .66, respectively, and in the control group .79 and .79. Pearson correlations of the SR with WT1 and WT2 were .41 and .49, respectively. Construct validity was examined by analyzing the performance as a function of task

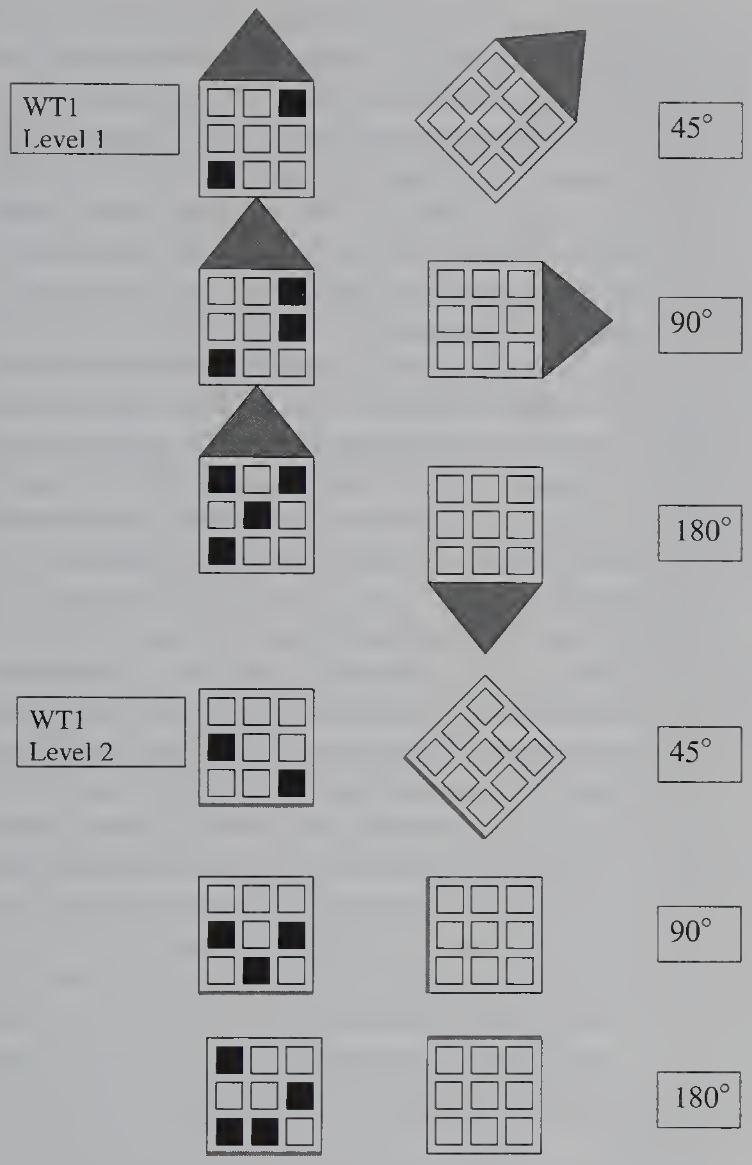


Figure 1. Mental rotation items from the Windows Test (WT).

characteristics. The findings of the pilot study showed clear linear decrease in performance as a function of degree of rotation, $F(2, 111) = 503.09$, $p < .001$, $\eta^2 = .90$; complexity level, $F(2, 111) = 28.53$, $p < .001$, $\eta^2 = .34$; symmetry, $F(1, 112) = 342.83$, $p < .001$, $\eta^2 = .75$; and test level (WT1, WT2), $F(1, 112) = 532.55$, $p < .001$, $\eta^2 = .83$. Findings on the initial composite WT scores, using the sample of the current study, also showed a significant performance decrease as a function of degrees of rotation, $F(2, 111) = 416.80$, $p < .001$, $\eta^2 = .88$, and test level, $F(1, 112) = 366.70$, $p < .001$, $\eta^2 = .77$.

Global–Local Processing Strategies. The GLPS (originally named Global–Local Judgment Tasks) was composed of 16 perceptual judgment tasks (stimulus cards) originally designed by Kimchi and Palmer (1982) for assessing visual–spatial processing styles among young children on a global–local dimension. Every stimulus card contained a standard figure at the top and two comparison figures at the bottom (see Figure 2). All the standard and comparison figures consisted of a global shape composed of local shapes. On each card, one of the comparison figures had the same global shape as the standard figure but comprised different local shapes; the other comparison figure had the same local shapes but was arranged in a different global shape. As can be seen in Figure 2, the global comparison figure is a square but comprised triangles. The local comparison figure, on the other hand, has the same local shapes (squares) but arranged into different global shape (triangle). For each stimulus card, children were asked to point out to one alternative out of two, which looked most like the standard, based on their first and immediate impression; clear instruction was provided that there are no right or wrong answers. The cards were presented one at a time in a random order, at a distance of approximately 60 cm from the child. To control for habitual orientation, half of the cards appeared with the global alternative on the right, while appearing on the left for the other half. A response was credited a score of 1 if the local comparison was selected and 2 if the global comparison was selected. The total score ranged from 16 to 32. Cronbach alpha reliabilities in a pilot study and in the current sample were .92 and .94, respectively.

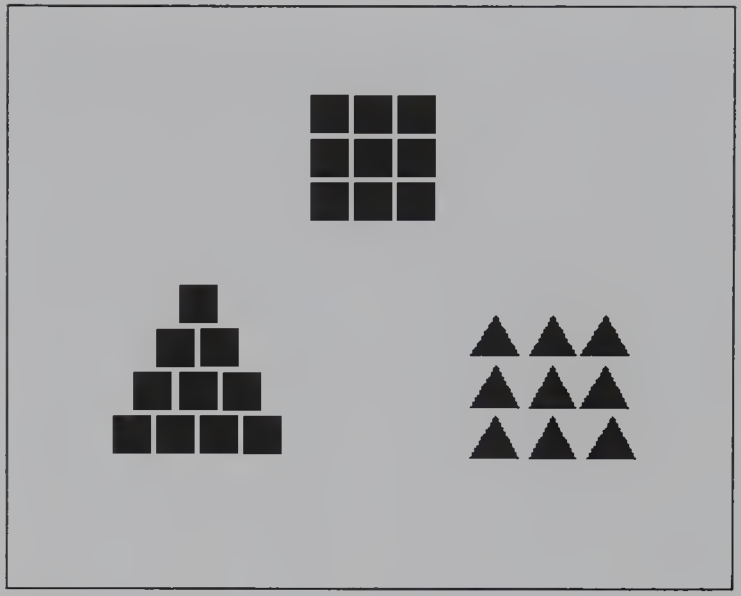


Figure 2. Example of a global–local task.

The "Spatial Sense" Intervention Program

The intervention program is based on Wheatley's (1996) "Quick Draws" activities, designed to promote spatial sense in the mathematics curriculum, with a focus on *representing and transforming spatial* *information,* according to the standards of the American National Council of Teachers of Mathematics.

The training program was composed of eight sessions, given once a week in a small group setting (3 boys and 3 girls). In each 45-min session, four flash cards were presented and discussed in the group. The activities were documented by the trainers and by children in special notebooks that served also for follow-up. Each lesson was composed of three stages. In the first stage, the trainer exposed flash cards to the children on a screen using an overhead projector. Each card was exposed three times, each time for 3 s. In the second stage, children were asked, after each series of card exposure, to draw what they saw. As the figure on the flash card is not observable while drawing, children must make their drawings using mental images they have constructed. The figures can be interpreted in many different ways (see Figure 3). In the third stage, the trainer guided the children, during group discussion, to look at the figures from different points of view. The children were encouraged to take their peers' perspectives and suggest the spatial imagery. The trainer allowed them to rotate the figures physically in order to demonstrate the image they perceived and drew. Typical descriptions given to the top picture shown in Figure 3, for example, were: a square with two trapezoids, two legs with shoes, a flower, a kite, a hat, a pyramid without top, a hallway, a room, a loudspeaker, back of a computer screen, and a popcorn box. Some of the descriptions were of a two-dimensional nature and some of three; some were constructed as an image of sections and others of joined segments. Typical descriptions given to the bottom picture shown in Figure 3, for example, were: hexagon with X, bow tie, two kites or diamonds, open envelope, a cube with some lines missing, and kind of a rollercoaster's seat.

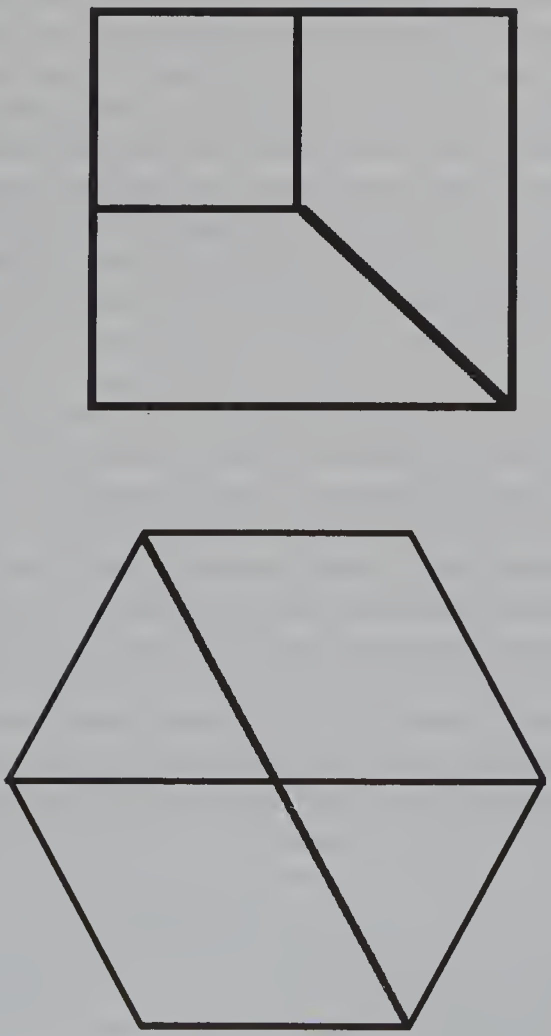


Figure 3. Training items from the "Spatial Sense" Intervention Program.

It should be emphasized that children were not directly trained to mentally rotate the figures but rather *to perceive the stimuli as a whole from different angles, retain them in their working memory, and flexibly conceptualize them as representing different objects.*

The original program is constructed for use with primary school children in a class situation. It includes 168 shapes arranged in seven levels. For the purposes of the current study, and based on a pilot study that verified their adequacy in terms of graphic-motor level, associative accessibility, and verbal possibilities, we selected 32 of the shapes found to be appropriate for use with first-grade children.

This choice of program was based on the studies of Wheatley and colleagues (Brown & Wheatley, 1997; Wheatley, 1990, 1991, 1997; Wheatley & Reynolds, 1999; Yackel & Wheatley, 1990), who showed that such activities improve children's ability to represent and transform spatial information. Analysis of the type of activities in the program shows that they help in expanding visuospatial working memory skills found to be crucial for mental rotation tasks (Coluccia & Louse, 2004). The choice of program is also appropriate to Tzuriel's (2000a, 2000b, 2001) approach, according to which concrete manipulations used with young children can serve as a "bridge" for complex mental manipulations.

Procedure

The procedure included three stages: preintervention, intervention, and postintervention. In the preintervention stage, children were individually tested on the SR test and WT (WT1, WT2), GLPS tasks, and Vocabulary and Concept Formation subtests, in that order. Graduate students administered the tests during 45 min in a quiet schoolroom.

Children in the experimental group participated in the intervention program for about a 3-month period. The control children participated in a substitute program for the same number of sessions, group structure, and setting. The control children were presented the same flash cards shown to the experimental group but without the short exposures, and were encouraged to copy the figures, individually, as accurately as possible. The control procedure was focused on pictographic and fine-motor skills but did not include teaching, group discussion, or guidance related to creation of visual imagery and spatial skills vital to mental rotations. Six trained graduate students who were not aware of the study's hypotheses administered the intervention and substitute programs. In the postintervention stage, 5 months after the preintervention phase, the SR test and WT (WT1, WT2) were administered again during one 30-min session. The tests in the postintervention phase were administered under the same conditions as the preintervention phase. Examiners who were not aware of the study's goals and the children's group assignment scored all tests.

Results

The Effects of Training Program on Gender Differences in Mental Rotation

The major objective of the current study was to investigate whether training in spatial abilities will moderate the effect of gender. In addition, we were interested in the relation between GLPS as a moderating factor of gender differences in spatial abilities and in the interaction effects of training with task characteristics (e.g., test level, degree of rotation) in moderating gender effects. Because of the complexity of the study's design, which includes five variables, we will report only interactions that involve gender, treatment, and time (pre- and postintervention). In Table 1, we present the pre- and postintervention scores of WT1, WT2, SR, and GLPS tests of boys and girls in the experimental and control groups. Preliminary correlation analyses of the study's variables were carried out, separately, for boys and girls (Appendix A) and for experimental and control groups (Appendix B). The findings indicate a similar correlation pattern for boys and girls and for experimental and control groups, except for the correlation of GLPS with SR Post, which was lower for girls than for boys.

Table 1

Pre- and Postintervention Windows Test (WT) and Spatial Relations (SR) Scores of Boys and Girls in the Experimental and Control Groups

		Experimental		Control	
Gender		Pre	Post	Pre	Post
WT Level 1					
Boys	*M*	46.63	50.07	45.89	47.14
	SD	4.41	3.06	3.98	3.77
Girls	*M*	43.93	49.83	41.86	43.89
	SD	5.38	2.63	5.49	5.62
WT Level 2					
Boys	*M*	39.00	45.00	37.07	38.79
	SD	5.99	4.58	6.72	5.34
Girls	*M*	34.77	43.17	34.14	35.11
	SD	6.07	4.43	5.35	5.98
SR test					
Boys	*M*	18.08	21.54	17.86	18.56
	SD	4.10	3.30	4.50	4.05
Girls	*M*	16.11	20.60	16.67	17.00
	SD	4.13	3.53	4.44	4.26

Analysis of WT scores. The effects of the training program were analyzed by a repeated measures nested design MANOVA of Treatment × Gender × Time × Test Level × Degree of Rotation ($2 \times 2 \times 2 \times 2 \times 3$), with treatment and gender as between-subjects factors and the last three variables as within-subjects factors; WT scores were the dependent variables.

The findings reveal a significant three-way interaction of Treatment × Gender × Time, $F(1, 112) = 4.73$, $p < .05$, $\eta_p^2 = .04$, as portrayed in Figure 4. Figure 4 shows that girls in the experimental group who scored much lower than boys did in the preintervention phase, improved their performance more than boys did and narrowed the gap in the postintervention phase. On the other hand, the girls in the control group scored equally lower than boys on both the pre- and postintervention phases. Post hoc analyses showed that in the experimental group, boys were higher than girls only in the preintervention phase, $t(1, 58) = 2.67$, $p < .01$, but not in the postintervention phase, $t(1, 58) = 1.11$, *ns*. In the control group, boys performed better than girls in both preintervention, $t(1, 54) = 2.55$, $p < .01$, and postintervention, $t(1, 54) = 2.83$, $p < .01$, phases. These findings support our main hypothesis by clearly showing that the intervention moderated and even canceled out the initial gender differences. Furthermore, in the postintervention phase, *the experimental girls performed significantly higher than the control boys*, $t(1, 56) = 3.79$, $p < .001$.

A significant three-way interaction of Treatment × Rotation × Time, $F(1, 112) = 13.51$, $p < .001$, $\eta_p^2 = .20$, shown in Figure 5, indicates that the gap between the experimental and control groups in the postintervention phase expands progressively with the increase in the degree of rotation. Post hoc analyses showed significant differences between the groups at 90°, $t(1, 114) = 6.15$, $p < .001$, and 180°, $t(1, 114) = 6.66$, $p < .001$, but not at 45°, $t(1, 114) = 1.29$, *ns*.

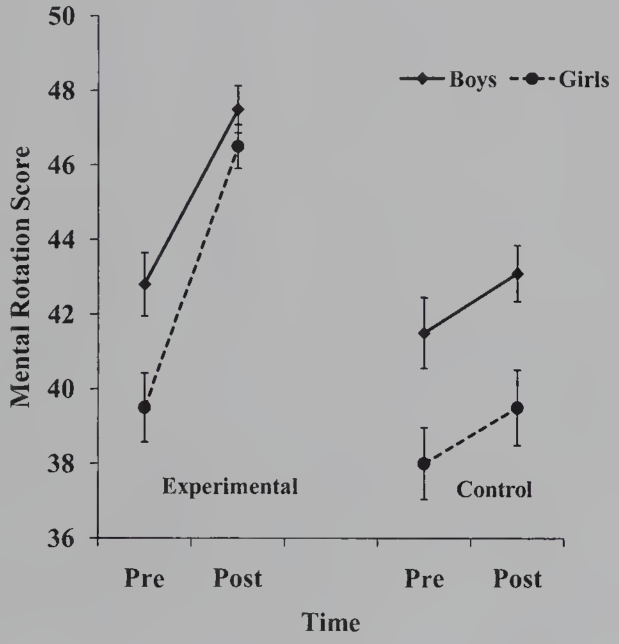


Figure 4. Pre- and postteaching performance on mental rotation among boys and girls in the experimental and control groups.

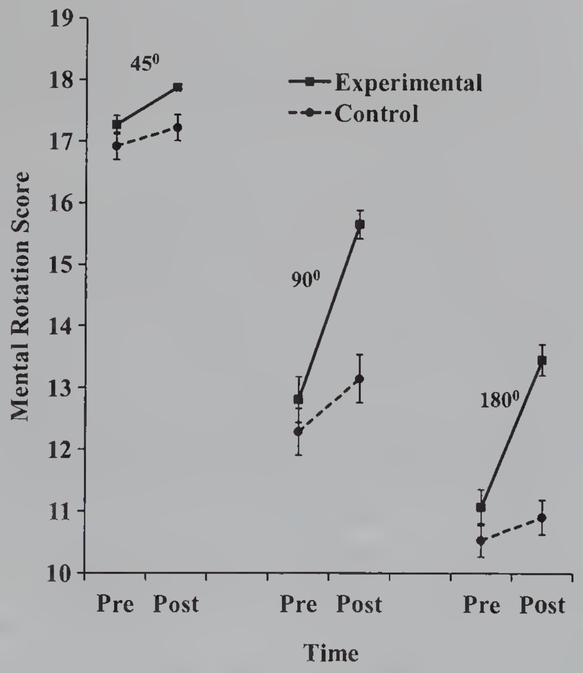


Figure 5. Pre- and postteaching performance on mental rotation in the experimental and control groups for each degree of rotation.

Another significant three-way interaction was of Treatment × Test Level × Time, $F(1, 112) = 8.69$, $p < .01$, $\eta_p^2 = .07$. This interaction, shown in Figure 6, indicates that the group differences in the postintervention phase were higher in Test Level 2 (WT2) than in Test Level 1 (WT1). Post hoc analyses showed that the experimental group scored significantly higher than control group in both Test Level 1, $t(1, 114) = 5.70$, $p < .001$, and Test Level 2, $t(1, 114) = 7.21$, $p < .001$. It is important to note that in Test Level 2, the experimental group showed much higher improvement, $t(1, 59) = 11.77$, $p < .001$, than the control group, $t(1, 55) = 2.31$, $p < .05$.

Analysis of SR test scores. We carried out separate analyses on the SR test as this test does not include task characteristics as the WT does. The means and standard deviations of boys and girls in the experimental and control groups are presented in Table 1. The effects of the training program were analyzed by a repeated measures MANOVA of Treatment × Gender × Time (2 × 2 × 2), with time as a within-subjects factor. The findings revealed a significant three-way interaction of Treatment × Gender × Time, $F(1, 112) = 71.80$, $p < .001$, $\eta_p^2 = .39$ (see Figure 7). As can be seen in Figure 7, gender differences in the experimental group decreased from pre- to postintervention, whereas in the control group, gender differences increased. Post hoc analyses showed that, in the experimental group, boys were higher than girls only in the preintervention, $t(1, 58) = 2.50$, $p < .01$, but not in the postintervention phase, $t(1, 58) = 1.19$, *ns*. In the control

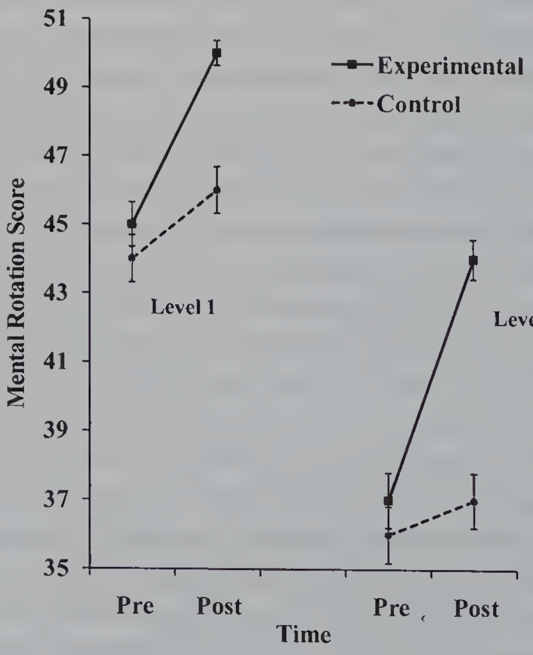


Figure 6. Pre- and postteaching performance on mental rotation in the experimental and control groups for each test level.

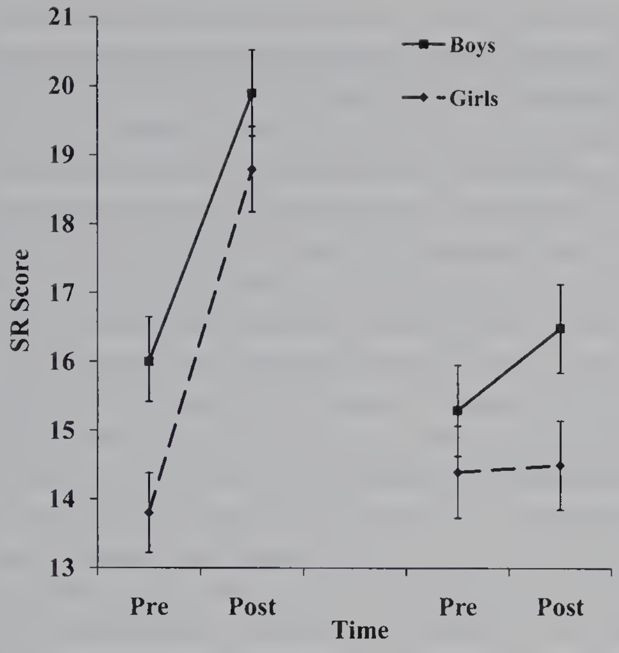


Figure 7. Pre- and postintervention Spatial Relations (SR) scores by treatment group and gender.

group, on the other hand, boys performed better than girls in both pre-, $t(1, 54) = 2.44$, $p < .01$, and postintervention, $t(1, 54) = 2.55$, $p < .01$, phases. Thus, the intervention moderated and even canceled out the initial gender differences. Furthermore, *in the postintervention phase the experimental girls performed significantly higher than the control boys*, $t(1, 56) = 2.68$, $p < .01$. The SR findings are very similar to those found for the WT and support out main hypothesis.

Effect-Size Analyses

In order to compare our findings with previous studies, we calculated the effect size of gender differences in the preintervention phase using Cohen's *d*. The results for gender, yielded an effect size of $d = .45$ ($r = .22$), $d = .69$ ($r = .32$), and $d = .60$ ($r = .28$) for the SR, WT1, and WT2, respectively. These effect sizes are very similar to the effect sizes found with older children and adults using spatial tasks like the Shepard Metzler mental rotation test (Voyer et al., 1995). It should be noted that the initial gender differences in our study are not likely to be attributed to general cognitive ability, as there were no gender differences on the WISC–R95 subscales of Vocabulary, $t(1, 114) = .64$, *ns*, and Concept Formation, $t(1, 114) = .30$, *ns*. Additional effect-size analyses were calculated for gender differences in the experimental and control groups before and after the intervention, using a composite score of WT (WT1 + WT2). The findings indicate that although in the experimental group the effect size decreased drastically from pre- (.71) to postintervention (.23), in the control group the effect size increased from pre- (.68) to postintervention (.73).

GLPS as a Moderating Factor Between Gender and Mental Rotation

The second objective of this study was to investigate whether the GLPS is a moderating factor that explains gender differences in mental rotation performance as well as the interactive effects of GLPS and task difficulty on mental rotation.

Prior to analyzing the effect of GLPS as a moderating factor of gender differences in mental rotation, we carried out two analyses to establish gender differences in GLPS. The first analysis was an ANOVA of Gender × Treatment (2 × 2) on GLPS. The ANOVA revealed a significant main effect of gender, $F(1, 112) = 25.76$, $p < .001$, $\eta^2 = .19$, indicating that boys were more global than were girls in their perceptual judgments, across treatment groups. Out of a possible total score of 16, the mean global choices selected were 10.14 ($SD = 2.15$) and 5.03 ($SD = 2.05$), for boys and girls, respectively.

The effect of the GLPS as a moderating variable was analyzed in a repeated measures MANCOVA of Gender × Treatment × Test Level × Degree of Rotation with GLPS as a covariate. The dependent variables were WT1 and WT2, which allowed the comparison of test levels and degrees of rotation. The findings revealed that (a) the gender main effect, which was significant in a previous MANOVA, $F(1, 112) = 12.80$, $p < .001$, $\eta_p^2 = .10$, was no longer significant in the current MANCOVA, $F(1, 111) = 1.75$, *ns*. Thus, once GLPS was included as a covariate, the gender main effect is not significant. (b) A significant interaction of GLPS × Degree of Rotation, $F(1, 111) = 5.08$, $p < .05$, $\eta^2 = .08$, indicates that the higher the degree of rotation, the higher is the contribution of a global orientation to mental rotation performance. As expected, the main effects of (c) test level, $F(1, 111) = 125.18$, $p < .001$, $\eta^2 = .53$; (d) degree of rotation, $F(1, 111) = 227.28$, $p < .001$, $\eta^2 = .81$; and (e) GLPS (covariate), $F(1, 111) = 31.51$, $p < .001$, $\eta^2 = .22$, were significant. These findings indicate that the higher the degree of rotation, test level, and the GLPS scores, the higher is the mental rotation score.

Further support for the relation between GLPS and mental rotation was indicated by correlation analysis. Pearson correlations between a combined preintervention WT score (WT1 + WT2) and GLPS score, showed positive significant correlations of .60 ($p < .001$) and .50 ($p < .001$) for the experimental and control groups, respectively. Similarly, the correlations between the GLPS and the initial SR scores were .30 ($p < .01$) and .28 ($p < .05$) for the experimental and control groups, respectively. These findings indicate that a global strategy was more efficient than a local strategy for solving mental rotation tasks.

Discussion

Data analysis of the pre- to postintervention scores showed clearly that the training program, focused on teaching of representation and transformation of spatial information, brought about a significant improvement in the spatial performance of young children, as well as a significant mitigation of the initial gap between boys and girls. The experimental group improved significantly more than the control group on both WT (Figure 4) and SR (Figure 7) tests. As the experimental and control groups showed high initial similarity on mental rotation scores, the observed improvement of the experimental group can be attributed mainly to the intervention effects. The slight improvement of the control group may indicate a self-practice effect caused by a repeated administration of the tasks. These findings differ from previous training studies, which indicate that the performance levels of males and females rise in parallel and maintain the gender gap (e.g., Baenninger & Newcombe, 1989; Levine et al., 1999).

Comparison of the pre- to postintervention improvement of boys and girls showed that girls' improvement in the experimental group was greater than that of boys and that the girls closed the initial gap. This finding was repeated in both the WT (Figure 4) and the SR (Figure 7) tests. Similar comparisons in the control group showed that the initial gender gap differences were not reduced from pre- to postintervention; on the WT, there were minor equal improvements among both boys and girls, and on the SR test, boys showed a slight improvement whereas girls showed no improvement. Moreover, in the postintervention phase, *girls in the experimental group showed higher performance than did boys in the control group; this improvement occurred on both tests*. Further support for the effectiveness of the intervention in reducing gender differences was evident in the effect-size findings. The effect size decreased sharply from pre- to postintervention in the experimental group but increased in the control group. A possible explanation for the uniqueness of our findings relates to the type of training given as well as to its intensity. It appears that self-training caused by repeated administration of the tasks, characterizing previous studies, was not alone sufficient to narrow the gap between boys and girls. As shown in previous studies (e.g., Marulis et al., 2007), males and females benefit equally from this type of self-training on spatial tasks. The findings of our experimental group revealed, on the other hand, that a relatively long-term *deliberated/instructive* training program based on teaching strategies for the transformation of visuospatial stimuli brought about substantial improvement in mental rotation of young children, particularly of girls who reached the same performance level as boys. The strategies focus on expanding working memory, perceiving spatial information in a global rather than analytic way, and flexibly conceptualizing spatial stimuli from different points of view. These findings are consistent with recent adult studies showing that females benefited more than males from active and guided training programs such as training with computer and video games, and that prior gender disparities in spatial ability, as measured by mental rotation tasks, were significantly reduced (Coluccia & Louse, 2004; Feng et al., 2007).

Support for our findings is found in Linn and Petersen's (1985) analysis of the processes that underlie the solution of mental rotation tasks. They suggested that a global-holistic or a Gestalt-like strategy is more effective than a part-by-part analytic approach. As the mental rotation tasks yield the largest and most robust gender differences, they raised the possibility that gender differences on these tasks are associated with the selection or efficient application of solution strategies. This argument is also supported by recent studies showing that boys tend more often than girls to adopt global-holistic strategy in processing visuospatial information since the early childhood years (e.g., Kramer et al., 1996; Lawton, Charleston, & Zieles, 1996; O'Laughlin & Brubaker, 1998).

Because mental rotation is considered to represent a high level of spatial cognition (Gardner, 1983) as well as the skill found to be most related to gender differences (Linn & Petersen, 1985; Voyer et al., 1995), our findings are of most importance. They suggest that early *deliberate/instructive* training programs might help to bridge the gender gap in spatial ability, whether the differences in ability are attributed to biological mechanisms or to environmental inputs. Moreover, our findings indicate that girls in the experimental group outperformed boys in the control group on both the SR and the WT tasks. These findings strengthen our claim that strategy-oriented intervention can overcome gender differences considered by many researchers to be biologically determined. Newcombe et al. (2002) noted that although the question of the origins that lead to the observed gender differences in spatial competence may be of scientific interest, *it is more important to determine whether intervention can reduce or eliminate them.* We would add that adequate intervention could reverse gender differences causing girls to perform higher than boys!

In the current research, first-grade boys showed higher initial performance than girls on a standardized SR test and even more on the WT, which was specifically developed to tap spatial skills involving mental rotation of young children. These initial gender differences are not likely due to intellectual ability, as boys and girls performed similarly on Vocabulary and Concept Formation subtests of the WISC–R or by SES level.

One of our main hypotheses was that processing strategies moderate and help to explain initial gender differences in mental rotation. By using Kimchi and Palmer's (1982) GLPS tasks, we first found that children who used a global-holistic strategy in their perceptual judgments were more likely to succeed in solving the mental rotation tasks, than children who used a local strategy: The GLPS scores were significantly and positively correlated with the SR and WT scores. The hypothesis is also strongly supported by two interrelated findings. The first is that once we introduced the GLPS variable as a covariate into the MANOVA (see previously mentioned MANOVA of Treatment × Gender × Time × Test Level × Degree of Rotation), the significant main effect of gender disappeared. The second is the finding showing that the GLPS score (introduced as a covariate) was significantly and positively related to the mental rotation scores. These findings support our assertion that a global-holistic strategy is more efficient than a local strategy in solving spatial tasks involving mental rotation.

Clinical observations of strategies used by boys and girls clearly show that boys used significantly more perceptual global judgments than girls did. Analysis of the GLPS scores indicated that boys tended to focus on the global shapes with more attention to the outer configurations, whereas girls tended to focus on the local elements in a more details-oriented fashion. This finding is consistent with Kramer et al. (1996), who used the same global–local paradigm with children between the ages of 4 and 12 years, and reported large gender differences with greater global bias in boys at all ages. These findings support our hypothesis that early gender differences in strategies used to process visuospatial information account for at least a portion of the gender differences in spatial performance as measured by mental rotation tasks. We would like to note that more evidence should be provided concerning the change in strategy usage from pre- to postintervention as well as the change in correlation patterns between mental rotation and processing strategies. Unfortunately, we did not gather information on the strategies used *during* testing; neither did we administer the GLPS test at the end of the intervention. In future research, we suggest use of the GLPS both before and after the intervention so that simultaneous comparisons can be carried out on mental rotation performance and processing strategies. We expect the intervention not only to affect both domains but also to observe a correlation between pre and postchanges in both domains. The findings showing that gender differences in mental rotation are moderated by differences in global–local strategies used to complement our expectations that a training program aimed at improving mental representation and transformation of visuospatial information would provide girls with opportunities to select and apply more efficient strategies for solving mental rotation tasks, leading to a reduction in gender differences.

In congruence with findings of meta-analytic studies (Linn & Petersen, 1985; Voyer et al., 1995), the magnitude of gender differences in the current study was influenced by task demands. Analysis of the interactive effects of treatment and task demands in WT revealed significant interactions of treatment with the degree of rotation (Figure 5) and test level (Figure 6). These interactions indicate that the higher the task difficulty, the higher was the impact of the intervention. On both variables, the experimental children showed significantly higher pre- to postintervention improvement on the difficult tasks than on easy tasks whereas the control group children showed consistent, slight and insignificant improvement.

Similar to the intervention effects on different levels of task difficulty, our findings indicate that the global strategy is more efficient in coping with tasks of high level than with low level of mental rotation. This was shown by the significant interaction of GLPS × Degree of Rotation and by the correlation analysis indicating higher correlations between WT and GLPS with increase in levels of rotation.

The origins that lead to the observed gender differences in strategies used for processing visuospatial information have not been clarified yet. On the one hand, evidence for biological mechanisms may derive from differences emerging by early childhood and remain relatively constant over the course of development (e.g., Kramer et al., 1996). On the other hand, these differences might be attributed to different experiences that are relevant to acquisition, selection, and application of solution strategies (Halpern, 2004; Linn & Petersen, 1985). Longitudinal and cross-sectional research showing parallel developmental trajectories of mental rotation and global–local strategies might help to establish the relation between the two domains.

In conclusion, given the fact that superior spatial skills are important in mathematical and scientific thinking (e.g., Clements & Battista, 1992; Gardner, 1983; Wheatley, 1990, 1991, 1996, 1997; Wheatley & Reynolds, 1999), it is crucial to implement intervention programs aimed at improving spatial skills and help to bridge the gap between males and females in these areas. In our study, mental rotation was found as a malleable domain that can be acquired through training. The findings indicate also that mental rotation is intimately related to global–local perception and that holistic strategies can help the individual in representing and transforming perceptual stimuli and directly affect mental rotation skills.

In view of the importance of spatial ability in many scientific domains and in view of the current results in supporting intervention for developing spatial skills, we recommend, together with others (e.g., Vasta et al., 1996; Webb et al., 2007), that more emphasis is needed to nurture such skills within the school curriculum. Training should include enhancing the processing strategies of visuospatial stimuli using computer technology (e.g., virtual reality) and targeted instructional approaches. Training should start as early as possible to prevent gender differences and to provide equal opportunities for girls to excel in spatial abilities that are required for success in scientific domains. Because girls have been shown to have fewer out-of-school spatial experiences than boys, many girls may never tap their potential to think spatially, unless spatial thinking is specifically introduced into the school curriculum. Nevertheless, further research is necessary to identify specific experiences that promote the development of spatial skills of young children.

References

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Role*, *20*, 327–344.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 12–23.

Brown, D., & Wheatley, G. (1997). Components of spatial reasoning. *Focus on Learning Problems in Mathematics*, *19*, 45–70.

Bryden, M. P., George, J., & Inch, R. (1990). Sex differences and the role of figural complexity in determining the rate of mental rotation. *Perceptual and Motor Skills*, *70*, 467–477.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, *31*, 697–705.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420–464). New York: Macmillan.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*, 329–340.

Cooper, L. A. (1976). Demonstration of mental analog of an external rotation. *Perception & Psychophysics*, *19*, 296–302.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex related differences in mathematics. *Intelligence*, *32*, 25–32.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850–855.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.

Geary, D. C. (2007). An evolutionary perspective on sex differences in mathematics and the sciences. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 173–188). Washington, DC: American Psychological Association.

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal*, *35*, 515–531.

Gur, R., & Gur, E. R. (2007). Neural substrates for sex differences in cognition. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 189–198). Washington, DC: American Psychological Association.

Haier, R. J. (2007). Brains, bias, and biology: Follow the data. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 113–120). Washington, DC: American Psychological Association.

Halpern, D. F. (2004). Cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science*, *13*, 135–139.

Halpern, D. F. (2007). Science sex and good sense: Why women are underrepresented in some areas of science and math. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 121–130). Washington, DC: American Psychological Association.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1–51.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications.* New York: Cambridge University Press.

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*, 250–261.

Johnson, A. M. (1990). Speed of mental rotation as a function of problem solving strategies. *Perceptual and Motor Skills*, *71*, 803–806.

Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development*, *58*, 725–740.

Kim, S. G., Ashe, K., Hendrich, J. M., Ellermann, H., Merkle, H., Ugurbil, K., et al. (1993). Functional magnetic resonance imaging of motor cortex: Hemispheric asymmetry and handedness. *Science*, *261*, 615–617.

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology*, *8*, 521–535.

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

Kramer, J. H., Leopard, J., Ellenberg, L., & Share, L. J. (1996). Developmental sex differences in global-local perceptual bias. *Neuropsychology*, *10*, 402–407.

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual and gender related differences in indoor way findings. *Environment and Behavior*, *28*, 204–219.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940–949.

Levine, S. C., Vasilyeva, M., Lourenco, F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*, 841–845.

Lidz, C. S., & Elliott, J. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. New York; Greenwich, CT: JAI Press.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.

Loring-Meier, S., & Halpern, D. (1999). Sex differences in visuospatial working memory: Components of cognitive processing. *Psychononic Bulletin and Review*, *6*, 464–471.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*, Stanford, CA: Stanford University Press.

Marulis, L. M., Lui, L. L., Warren, C. M., Uttal, D. H., & Newcombe, N. S. (2007, April). Effects of training or experience on spatial cognition in children and adults: A meta-analysis. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337–341.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066.

Newcombe, N. S. (2007). Taking science seriously: Straight thinking about spatial sex differences. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 69–78). Washington, DC: American Psychological Association.

Newcombe, N. S., Mathason, L., & Terlecki, M. (2002). Maximization of spatial competence: More important than finding the cause of sex differences. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society and behavior: The development of sex differences in cognition* (pp. 183–206). Westport, CT: Ablex.

Nordvik, H., & Amponsah, B. (1998). Gender differences in spatial abilities and spatial activity among university students in an egalitarian educational system. *Sex Roles*, *38*, 1009–1023.

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences*, *24*, 595–601.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*, 1067–1070.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Snow, R. E., & Yalow, E. (1982) Education and intelligence. In R. J. Sternberg (Ed.), *A handbook of human intelligence* (pp. 493–585). Cambridge, UK: Cambridge University Press.

Sorby, S. A., & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students. *Journal of Engineering Education*, *89*, 301–307.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing*. New York: Cambridge University Press.

Thurstone, L. L., & Thurstone, T. G. (1965). *Primary Mental Abilities tests*. Chicago: Science Research.

Tzuriel, D. (1995). *The Cognitive Modifiability Battery (CMB). Assessment and intervention: User's manual*. Tel Aviv, Israel: School of Education, Bar-Ilan University.

Tzuriel, D. (2000a). The Cognitive Modifiability Battery: Assessment and intervention. In C. S. Lidz & J. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp. 375–406). New York: Elsevier Science.

Tzuriel, D. (2000b). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, *12*, 385–435.

Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Kluwer Academic/Plenum.

Tzuriel, D., & Egozi, G. (2006). Dynamic assessment of young children's spatial abilities: The effects of gender and task characteristics. *Journal of Cognitive Education and Psychology*, *6*, 218–247.

Vasta, R., Knott, J. A., & Gaze, C. E. (1996). Can spatial training erase the gender differences on the water-level task? *Psychology of Women Quarterly*, *20*, 549–568.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*, 397–420.

Wheatley, G. (1990). Spatial sense and mathematics learning. *Arithmetic Teacher*, *37*, 10–11.

Wheatley, G. (1991). The role of imagery in mathematics learning. *Arithmetic Teacher*, *39*, 34–36.

Wheatley, G. (1996). *Quick draw: Developing spatial sense in mathematics*. Tallahassee: Florida Department of Education.

Wheatley, G. (1997). Reasoning with images in mathematical activity. In L. D. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 281–298). Mahwah, NJ: Erlbaum.

Wheatley, G., & Reynolds, A. (1999). Image maker: Developing spatial sense. *Teaching Children Mathematics*, *5*, 374–378.

Yackel, E., & Wheatley, G. (1990). Promoting visual imagery in young children. *Arithmetic Teacher*, *37*, 52–58.

Appendix A

Matrix of Correlations of the Study's Variables for Boys and Girls

	WT Pre	WT Post	SR Pre	SR Post	GLPS
WT Pre	—	.72**	.34**	.45**	.46**
WT Post	_.70**_	—	.32*	.57**	.39**
SR Pre	_.46**_	_.28*_	—	.71**	.39**
SR Post	_.40**_	_.56**_	_.54**_	—	.38**
GLPS	_.50**_	_.40**_	_.28*_	_.16_	—

Note. Plain font = boys; italic font = girls. WT = Windows Test, general score; SR = Spatial Relations test; GLPS = Global–Local Processing Strategies.
$^{*}p < .05$. $^{**}p < .01$.

Appendix B

Matrix of Correlations of the Study's Variables for Experimental and Control Groups

	WT Pre	WT Pre	SR Pre	SR Post	GLPS
WT Pre	—	.72**	.48**	.46**	.60**
WT Post	_.86**_	—	.40*	.50**	.43**
SR Pre	_.41**_	_.34*_	—	.75**	.30**
SR Post	_.44**_	_.37**_	_.76**_	—	.34**
GLPS	_.50**_	_.52**_	_.28*_	_.34*_	—

Note. Plain font = experimental; italic font = control. WT = Windows Test, general score; SR = Spatial Relations test; GLPS = Global–Local Processing Strategies.
$^{*}p < .05$. $^{**}p < .01$.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1431–1444

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents

Tiffany Yip
Fordham University

Eleanor K. Seaton
University of North Carolina, Chapel Hill

Robert M. Sellers
University of Michigan

Among 224 African American adolescents (mean age = 14), the associations between interracial and intraracial contact and school-level diversity on changes in racial identity over a 3-year period were examined. Youths were determined to be diffused, foreclosed, moratorium, or achieved, and change or stability in identity status was examined. Contact with Black students, Black friends, and White friends predicted change in identity status. Furthermore, in racially diverse schools, having more Black friends was associated with identity stability. Students reporting low contact with Black students in racially diverse schools were more likely to report identity change if they had few Black friends. In students reporting high contact with Blacks in predominantly White schools, their identity was less likely to change for students with fewer White friends.

Intergroup contact theory suggests that, under certain conditions, the more intergroup contact individuals have, the more likely they are to feel positive toward out-group members (Pettigrew, 1998). Indeed, research has shown that positive interactions with different-race individuals have been associated with outcomes such as reduction in prejudice (Molina & Wittig, 2006), increases in a sense of common identity (Nier et al., 2001), and personal closeness (Tropp, 2007). In a meta-analytic review of studies that have examined the contact hypothesis, Pettigrew and Tropp (2006) reported that intergroup contact has a positive impact on reducing prejudice for members of the general outgroup and not just specific members with whom an individual has had interactions. However, as Pettigrew (1998) writes, "optimal intergroup contact provides insight about ingroups as well as outgroups" (p. 72). Indeed, research suggests that in-group and out-group contact influences the development of racial identity (Alba & Nee, 2003; Portes & Rumbaut, 2001). The current study will address two gaps in the current literature on the effects of contact on adolescent development. First, the study will consider the influences of both intergroup and intragroup peer contact and friendships on changes in youth's feelings about their own racial identity over time. Finally, this study will consider the opportunities that youth have for interracial and intraracial contact and friendships by examining how contact interacts with school-level diversity to influence racial identity changes.

Racial Identity Development

Many scholars have discussed the importance of identity exploration and construction during adolescence (Erikson, 1968; Marcia, 1980; Phinney, 1993). Indeed, identity status models outline how the combination of identity exploration and commitment change as adolescents learn more about their racial group. Specifically, Marcia (1966) and Phinney (1992) offer a model of identity development where adolescents can be in one of four statuses: neither explored nor committed to their identity (diffused), committed to an identity without exploring (foreclosed), explored but have not committed (moratorium), or have both explored and committed to an identity (achieved). Inherent

---

This research was supported by grants from the National Institute of Mental Health (NIMH 5 R01 MH061967-03) and the National Science Foundation (BCS-9986101) awarded to the third author.

Correspondence concerning this article should be addressed to Tiffany Yip, Department of Psychology, Fordham University, 332 Dealy Hall, Bronx, NY 10458. Electronic mail may be sent to tyip@fordham.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0009

in this model is a pattern of linear progression from less to more advanced stages. Recently, empirical work has lent some support to this theory finding that adolescents were more likely to report being in the less advanced statuses when compared with older adults (Yip, Seaton, & Sellers, 2006). Other research finds that even from one year to the next, adolescents report moving from one identity status to another (Seaton, Scottham, & Sellers, 2006). Thus, racial identity appears to be changing for minority youth during the adolescent period, but the empirical literature is less clear about antecedents to change in racial identity status.

Intergroup Contact and Racial Identity

There are many possibilities for what might precede changes in identity. One such consideration is social interactions with other-race peers. In one of the few studies examining the association between racial identity and intergroup contact, Lee, Noh, Yoo, and Doh (2007) found that among Koreans living in China, those who reported more social interactions with Chinese reported lower levels of ethnic identity when discrimination was frequent. In a study of African, Asian, and European American adolescents, Hamm (2000) found that only for European Americans did having cross-race friendships have an association with ethnic identity. In particular, it was observed that having a cross-race friendship was associated with higher scores on importance and positive regard for one's ethnicity. In a study of Latino adolescents, attending a predominantly non-Latino school was associated with stronger ethnic identity (Umana-Taylor, 2004). Together, these studies suggest intergroup contact may have important implications for identity development. However, as none of these studies employed longitudinal data, the nature of the relation between contact and racial identity is unclear.

Intragroup Contact and Racial Identity

Although the contact hypothesis focuses primarily on the effects of intergroup contact, research has also examined the very important influence of intragroup interactions for youth development. For example, in a study of African American adolescents, Rowley, Burchinal, Roberts, and Zeisel (2008) found that same-race friendships were associated with increased reports of racial discrimination. Although same-race friends were not linked to racial identity explicitly in that study, recent research suggests that discrimination is a form of social interaction that may then lead to the exploration of and commitment to one's identity (Seaton, Yip, & Sellers, 2009). Specifically, theory suggests that experiences of racial discrimination may serve as encounters that cause individuals to rethink the role of race in their lives, a process that can occur anywhere in the developmental life span (Cross & Fhagen-Smith, 2001). In research that considered ethnic identity as an outcome, researchers found that having a stronger sense of ethnic identification was associated with more in-group peer interaction (Phinney, Romero, Nava, & Huang, 2001). In retrospective research on Black college students who are attending predominantly White high schools, it was observed that having positive same-race peer relationships was associated with more resolution in racial identity formation (Tatum, 2004). Therefore, it seems that having more same-race friends may be associated with changes in racial identity status over time. Taken together, the current empirical literature suggests that both intergroup and intragroup contact could be related to changes in racial identity.

School Diversity

Implicitly, intergroup contact theory suggests that individuals should have opportunity to interact with out-group members in everyday contexts. Indeed, scholars have discussed the importance of opportunity for contact in the context of macro-level variables such as diversity, in their discussion of intergroup contact theory (de Souza Briggs, 2007; Pettigrew, 2006). As the primary context in which adolescents have contact with peers is school, this study considers school-level diversity as an indicator of the opportunities that youths have to interact with same- and different-race peers. Schools that are characterized by high diversity afford youths increased opportunity to interact with same- *and* different-race peers. Schools with low diversity, however, may only offer youths the opportunity to interact with same- *or* different-race peers, but not both. Therefore, the current study considers school-level diversity as an important third variable to consider in examining the association between contact and changes in racial identity.

Furthermore, even with the opportunity to interact with same- and different-race peers, adolescents may not actually interact with them. For one, there may be institutional barriers; for example, classes may be organized in a manner that is not conducive to interactions between students in different academic tracks (Kubitschek & Hallinan, 1998; Mickelson &

Heath, 1999). Specifically, research suggests that academic tracks often differ by race (Schwarzwald & Cohen, 1982). Beyond academic divisions, in social interactions, it has also been observed that youths may self-segregate and befriend only same-race peers (Tatum, 1997). For example, a study of Dutch adolescents found that despite attending diverse schools where at least half of the students were not Dutch, 16% of the Dutch students reported having no contact with non-Dutch peers (Masson & Verkuyten, 1993). Outside of school, where students presumably have more choice about whether or not to engage in intergroup interactions, 45% of students reported little or no cross-ethnic interactions. In this same study, students reporting a more positive attitude about their racial group also reported less intergroup contact. In a study of Black and White fourth grade students, researchers found that students attending an integrated school were much more likely to mention race and ethnicity in discussions about self-concept than children attending predominantly Black or White schools (Dutton, Singer, & Devlin, 1998). As such, it seems important to consider not only youths' opportunities to interact with same- and different-race peers, but also behavioral indicators of whether or not youth actually make friends with same- and different-race peers.

Friendships

Therefore, another approach to evaluating the contact hypothesis (Pettigrew, 2007) includes research in the area of cross-race friendships. As an operationalization of the contact hypothesis, studies of cross-race friendships have lent support to the basic tenets of the theory. For example, research examining school-level racial diversity on friendship choices among high school students finds that the higher the proportion of same-race peers in a student's school, the more likely he or she will be to have same-race friends (Kao & Joyner, 2006; Kao & Vaquera, 2006). Similarly, other research suggests that cross-race friendships are more likely in more racially diverse schools (Quillian & Campbell, 2003). Patterns for ethnicity are analogous, such that attending a school with more same-ethnicity peers also makes it more likely that students will report having more same-ethnicity friends (Kao & Joyner, 2006).

From a probabilistic point of view, it seems logical that the more opportunity one has to form cross-race friendships, the more likely the friendships will occur. However, research suggests that actual behavior patterns are less predictable. Research on Black and White school children has found that students in integrated and predominantly White schools were more likely to report having cross-race friendships (Dutton et al., 1998). In a study of Black and White students attending an integrated junior high school, most students reported having a close other-race friend; however, only 28% of the sample reported having interactions with other-race friends off campus (DuBois & Hirsch, 1990). At the same time, research on adolescents has observed when one's group represents a small minority in an otherwise homogenous school, the probability of same-race friendship increases (Quillian & Campbell, 2003). In a study with adults, using multilevel models that estimated the probability of establishing cross-race friendships given the relative proportions of different ethnic groups in one's community, de Souza Briggs (2007) analyzed the friendship patterns of over 20,000 adults in the United States. These data suggested that the single greatest predictor of whether or not a European American adult reported having a cross-race friendship was the region of the country in which the individual resided. Even so, the actual rates of cross-race friendships fall short of the predicted probability (as estimated by a random choice model given the racial distribution of one's local area) for all ethnic groups (de Souza Briggs, 2007). In other words, even though living in a more diverse area provided increased opportunity to form cross-race friendships, individual differences still account for whether or not people actually formed cross-race friendships.

The current study will examine the prospective effects of intergroup and intragroup contact and friendships on changes in racial identity status in a 3-year longitudinal sample of African American adolescents living in a predominantly White Midwestern town. Interracial and intraracial contact is operationalized in two ways: (a) as students' reports of the proportion of other Black youth in their classes and clubs and (b) as friendships, the number of close White and close Black friends each student reports. We hypothesize that interracial contact and intraracial contact may or may not have independent effects on whether youths' racial identities change or remain stable. Instead, we do expect that contact and context will interact to determine whether there are changes in racial identity over time. For the purposes of this study, context is operationalized as diversity using school-level information about the racial composition of each student's school.

Method

Participants

These data are drawn from a 3-year longitudinal study of racial identity development, racial socialization, and psychological adjustment among African American adolescents residing in the Midwest. About 15% of residents in the school district reported being African American, whereas 54% reported being of European American descent. There were more females than males in the study (60%) and the median family income was reported by parents to be between $40,000 and $49,000. Although the median income for African American families in this community is greater than the national average range ($30,000–$39,000), there is variation for the participants' families.

Students began the study either in the 7th, 8th, 9th, or 10th grade (mean age = 13.83 years old, $SD = 1.11$), and were administered surveys once a year for 3 years. Students attended 10 schools in the same school district (6 middle and 4 high schools). At all three time points, data were collected from both an adolescent and his or her primary caregiver. Five hundred and forty-six students completed surveys during the 1st year of the study. In Year 2, 379 of those students returned. A total of 260 adolescents completed all three time points. With the exception of one school, all schools were predominantly White; therefore, for ease and consistency of interpretation, the predominantly Black school was omitted from the analyses presented here. As such, the current study includes the 224 adolescents who had complete data at all three time points. To address whether the data were missing at random, chi-square and analyses of variance (ANOVAs) compared adolescents with complete data across the three time points with adolescents who were omitted from the current study for not having data at all three times points. Examining gender and racial identity variables, there were no significant differences between adolescents who completed all three waves of the study compared with those who only completed one or two waves: gender, $\chi^2(2) = 1.08$, *ns*; exploration, $F(2, 543) = 0.28$, *ns*; and commitment, $F(2, 543) = 2.49$, *ns*.

Procedure

The specific procedures for the longitudinal study have been described elsewhere (Seaton et al., 2009). All adolescents identified as African American by the school district were contacted by phone or mail to participate in the study. At each survey administration, adolescents were asked to sign assent forms. Adolescents completed pencil-and-paper surveys in small groups after school and sessions lasted between 45 and 75 min. Students were compensated with $30, $40, and $50 gift certificates to the local mall for each year of the study, respectively.

Measures

Racial identity. The Multigroup Ethnic Identity Measure (MEIM) comprised four subscales: affirmation and belonging, identity achievement, racial behaviors, and other-group orientation (Phinney, 1992). The identity achievement subscale, which measured exploration and commitment, was used in the present study. The scale consists of items with responses ranging from 1 (*strongly disagree*) to 4 (*strongly agree*), such that higher scores represent strong endorsement of their respective constructs. Previous research using the MEIM reported a two-factor solution consisting of affirmation and belonging (affirmation, belonging, and commitment items) and identity search (exploration and ethnic behavior items; Phinney, 1992; Roberts et al., 1999). However, previous work has not exclusively focused on the achievement subscale, which measures exploration and commitment (Phinney, 1992). As the goal of the present study was to create the identity statuses from exploration and commitment levels, it was appropriate to separate the achievement subscale into exploration and commitment dimensions.

A confirmatory factor analysis was conducted to discern whether the achievement subscale consisted of two oblique latent factors with exploration and commitment items at all three time periods. The results indicate an average fit to the data for Time 1 ($\chi^2 = 25.19$, $df = 8$, $\chi^2/df = 3.15$, Tucker-Lewis index [TLI] = .82, comparative fit index [CFI] = .93, root mean square error of approximation [RMSEA] = .10), Time 2 ($\chi^2 = 18.49$, $df = 8$, $\chi^2/df = 2.31$, TLI = .91, CFI = .96, RMSEA = .08), and Time 3 ($\chi^2 = 47.99$, $df = 8$, $\chi^2/df = 5.99$, TLI = .62, CFI = .86, RMSEA = .15), providing an empirical justification for separating the achievement subscale of the MEIM into exploration and commitment indices.

As analyses were conducted over a 3-year time frame to measure change in racial identity, it was important to establish that observed change is due to differences in youths' reports and not measurement variance. In order to examine invariance, differences in the chi-square statistic were examined over time for the unconstrained model, the constrained measurement weights model, and the constrained measurement intercepts model. Differences in the chi-square statistic involve comparing the change in χ^2 with a critical ratio. The test comparing the unconstrained model with the constrained measurement weights model was not significant, $\Delta\chi^2 = 7.05$, with the critical ratio equivalent to 15.51. The test comparing the unconstrained model with the constrained measurement intercepts model was not significant, $\Delta\chi^2 = 28.75$, with the critical ratio equivalent to 31.41. Thus, the results suggest that the confirmatory factor analytic models are invariant across time.

The resulting exploration subscale consists of four items rated on a 4-point Likert scale, which range from strongly disagree to strongly agree (Time 1: $M = 2.87$, $SD = 0.63$; Time 2: $M = 2.82$, $SD = 0.57$, Time 3: $M = 2.79$, $SD = 0.64$), and a sample item includes ''I think a lot about what being Black means for my life'' (Time 1: $\alpha = .66$; Time 2: $\alpha = .63$; Time: 3 $\alpha = .64$). The commitment subscale has two items rated on a 4-point scale (Time 1: $M = 3.09$, $SD = 0.70$; Time 2: $M = 3.02$, $SD = 0.72$, Time 3: $M = 2.98$, $SD = 0.72$) and a sample item includes ''I understand pretty well what being Black means to me'' (Time 1: $\alpha = .68$; Time 2: $\alpha = .69$; Time 3: $\alpha = .71$).

Because the Cronbach's alphas for the exploration and commitment variables included significantly fewer items than the full-item MEIM, the Spearman–Brown formula was used to calculate the estimated Cronbach's alphas if the subscales included 12 items (equivalent to the full-item MEIM). The estimated Cronbach's alphas are: exploration (Time 1: $\alpha = .83$; Time 2: $\alpha = .83$; Time 3: $\alpha = .85$) and commitment (Time 1: $\alpha = .83$; Time 2: $\alpha = .84$; Time 3: $\alpha = .82$). These coefficients are comparable with existing research employing the 12-item MEIM; for example, $\alpha = .82$ (Roberts et al., 1999), $\alpha = .75$ (McMahon & Watts, 2002), and $\alpha = .82$ (Bracey, Bamaca, & Umana-Taylor, 2004). Scores for each subscale were developed by averaging across the items, and higher scores are indicative of high levels of exploration or commitment.

Interracial contact. Two questions were asked: ''Which of the following best describes the racial makeup of the students in most of your classes?'' and ''Which of the following best describes the racial makeup of the people in most of the clubs, teams, or other organizations you are currently involved in?'' Responses included 1 (*almost all White students*), 2 (*more White than Black students*), 3 (*same number of Black and White students*), 4 (*more Black than White students*), and 5 (*almost all Black students*). These two items were averaged together to estimate contact ($M = 2.37$, $SD = 0.92$). Higher scores indicate more contact with same-race, Black youths.

Friendships. As another measure of contact (interracial and intraracial), we assessed youths' reported friendships. ''How many of your close friends are Black?'' assessed intraracial close friendships. An analogous question, ''How many of your close friends are White?'' assessed interracial close friendships. Responses included 1 (*none*), 2 (*a few*), 3 (*half*), 4 (*most*), and 5 (*all*). To keep interracial and interracial friendships distinct, these items were kept separate and higher scores represent more close friendships with Black and White students ($M = 3.07$, $SD = 0.78$; $M = 2.60$, $SD = 0.75$, respectively).

School diversity index. The racial diversity of each school was obtained with archival data from the school district. The racial diversity of each student's school was computed based on information about the percentage of African American, Asian, White, Hispanic, Middle Eastern, Native American, multiethnic, and other students. Employing methods described in Juvonen, Nishina, and Graham (2006), the racial diversity of each student's school was computed based on information about the percentage of students in each racial group. Using information about the number of different racial/ethnic groups (g) and the proportion of individuals (p) who are members of each group (i), the index (D_C) provides an estimate of the relative probability that two randomly selected students are from different racial/ethnic groups:

$$D_C = 1 - \sum_{i-1}^{g} p_i^2.$$

Higher scores (range from 0 to 1) indicate more diversity (i.e., presence of other racial minorities), whereas lower scores indicate less diversity (i.e., predominantly White) and indicators were computed for all three time points (Time 1: $M = .57$, $SD = 0.09$; Time 2: $M = .57$, $SD = 0.08$; Time 3: $M = .58$, $SD = 0.08$).

Results

Correlations between the primary study variables are presented in Table 1. Among the Time 1 variables, reporting having more contact with Black students in classes and clubs was positively associated with reporting having more close Black friends, fewer close White friends, and higher

Table 1

Correlation Between the Time 1, Time 2, and Time 3 Study Variables

	Time 1				Time 2				Time 3			
	1	2	3	4	1	2	3	4	1	2	3	4
Time 1												
1. Composition of classrooms and clubs	—	.27**	−.16**	.18*	.38**	.25**	−.27**	.02	.36**	.28**	−.26**	.05
2. Close Black friends		—	−.67**	.11	.25**	.64**	−.54**	.00	.27**	.64**	−.53**	−.03
3. Close White friends			—	−.07	−.11	−.59**	.66**	.03	−.19**	−.56**	.55**	.04
4. School diversity				—	.22**	.18**	−.19*	.82**	.07	.17*	−.22**	.75**
Time 2												
1. Composition of classrooms and clubs					—	.25**	−.19**	.19**	.37**	.31**	−.19**	.11
2. Close Black friends						—	−.64**	.05	.28**	.62**	−.49**	.07
3. Close White friends							—	−.08	−.26**	−.55**	.64**	−.02
4. School diversity								—	−.03	.07	−.11	.84**
Time 3												
1. Composition of classrooms and clubs									—	.31**	−.32**	−.06
2. Close Black friends										—	−.67**	.07
3. Close White friends											—	−.09
4. School diversity												—

*$p < .05$. **$p < .01$.

school diversity. Reporting more close Black friends was associated with having fewer close White friends. Similar patterns were observed among the Time 2 variables. Among the Time 3 variables, one difference is noted: The composition of classes and clubs is no longer associated with the school diversity index.

Examining the correlations across adjacent time points, Time 1 composition of classes and clubs is positively correlated with reports at Time 2. Time 1 composition of classes and clubs is also positively associated with having more close Black friends but negatively associated with having more close White friends. Reporting more close Black friends at Time 1 was positively associated with having more contact with Black students, more close Black friends, and fewer White close friends at Time 2. The number of close White friends reported at Time 1 was negatively associated with close Black friends at Time 2, but positively associated with close White friends at Time 2. School diversity at Time 1 was associated with more contact with Blacks; more close Black friends, and school diversity at Time 2. However, school diversity at Time 1 was negatively associated with close White friends at Time 2.

Examination of the Time 2 and Time 3 adjacent time points suggest that there was stability in composition of classes and clubs. There was also a positive association between composition and close Black friends, but a negative association with close White friends. Consistent with previous time points, there was a negative association between reports of close White friends and close Black friends between Time 2 and Time 3, but stability in reports of the same racial groups over time. Finally, stability was observed for reports of school diversity at Time 2 and Time 3. Looking over a 2-year period, having more Black students in classes and clubs at Time 1 was positively associated with the same variable at Time 3. And as with previous time points, Time 1 contact was positively associated with having more close Black friends and negatively associated with having close White friends. Having many close Black friends at Time 1 was associated with having many close Black friends at Time 3, fewer close White friends, and reporting more Blacks in classes and clubs. Again, stability was observed for not only having close White friends from Time 1 to Time 3, but also associated with having fewer close Black friends and reporting fewer Black students in classes and clubs. The school diversity index at Time 1 was positively associated with having more close Black friends at Time 3, as well as school diversity at Time 3. School diversity at Time 1 was negatively associated with reports of close White friends at Time 3. As the school diversity index over the three time points was so highly correlated, a mean was computed of the three indices and this index was used for the analyses that follow.

Racial Identity Statuses

Before examining change in racial identity statuses, we first determined which racial identity clusters were present in the data. Consistent with racial identity status theory and with previous research (Yip et al., 2006), we expected to find four identity statuses that reflect the diffused, foreclosed, moratorium, and achieved profiles. *K* means cluster analyses were utilized to identify homogenous racial identity statuses among standardized exploration and commitment variables. *K* means analyses are the most appropriate technique when there is a theoretical rationale for a specific number of clusters (Hair & Black, 2000). The specific statistical techniques used to create the four identity status groups have been described elsewhere (Yip et al., 2006). Consistent with Phinney's (1989) identity formation model, the data suggested the presence of four distinct identity profiles. The first profile had significantly lower scores on the exploration and commitment subscales and was classified as diffuse (Time 1: $n = 21$; Time 2: $n = 31$; Time 3: $n = 36$). A second cluster reporting lower scores on the exploration subscale and higher scores on the commitment subscale was classified as foreclosed (Time 1: $n = 48$; Time 2: $n = 55$; Time 3: $n = 28$). A third cluster fits the description for moratorium, with higher scores on the exploration subscale and lower scores on the commitment subscale (Time 1: $n = 65$; Time 2: $n = 74$; Time 3: $n = 80$). The final cluster was classified as achieved, with higher scores on both the exploration and commitment subscales (Time 1: $n = 77$; Time 2: $n = 51$; Time 3: $n = 67$).

Interracial/Intraracial Contact and School Diversity

The racial identity status variable assessed whether individuals changed over time. Individuals who remained in the same status over time were labeled as *constant*, whereas individuals who changed identity status groups over time were labeled as *change*. Employing these coding methods, we observed that between Time 1 and Time 2, 42% of the sample reported no change in their racial identity status whereas the remaining 58% reported a change in their identity status. Between Time 2 and Time 3, 36% of the sample reported remaining in the same identity status whereas 64% did not. Finally, over the 2-year period between Time 1 and Time 3, 38% of students reported staying in the same identity status whereas 62% moving from one status to another.

The next step was to examine how interracial and intraracial contact prospectively predicted the observed changes in identity status. As such, binomial logistic regression analyses were conducted where identity change from Time 1 to Time 2 was coded as 1, and identity stability from Time 1 to Time 2 was coded as 0. Student age and gender were included as covariates. School diversity (mean of Time 1, 2, 3), number of close White friends (Time 1), number of close Black friends (Time 1), contact with Black students in classrooms and clubs (Time 1), the three possible interactions of school diversity with each of the three other independent variables, and the two possible interactions between school diversity, contact, and number of close Black and White friends were included as predictors (Table 2). A main effect of contact with Blacks in classrooms and clubs was observed such that more contact was associated with an increased likelihood of change in racial identity status. In addition, a three-way interaction between contact, diversity, and number of close Black friends was found, which is illustrated jointly in Figures 1 and 2. To graph the three-way interaction, a median split was computed for contact with Black students (Figure 1: low contact with Blacks; Figure 2: high contact with Blacks). Among students who reported having low contact with Black students (Figure 1), those with few close Black friends who reported attending schools with higher diversity (i.e., presence of other minorities) reported being more likely to change their identity status from Time 1 to Time 2. On the other hand, among students who reported having high contact with Black students

Table 2

Logistic Regression Predicting Probability of Identity Status Change From Time 1 to Time 2

	B	*SE*
Constant	0.66	1.93
Age	0.03	0.13
Gender	−0.50	0.31
School diversity (mean Time 1, 2, 3)	1.54	1.57
Close White friend—Time 1	0.11	0.25
Close Black friend—Time 1	−0.04	0.22
Contact—Time 1	0.50**	0.19
School Diversity × Close White Friend—Time 1	1.00	2.65
School Diversity × Close Black Friend—Time 1	−0.46	2.39
School Diversity × Contact—Time 1	1.08	1.87
School Diversity × Contact × White Friend—Time 1	−5.62	3.29
School Diversity × Contact × Black Friend—Time 1	−6.49*	3.05

*$p < .05$. **$p < .01$.

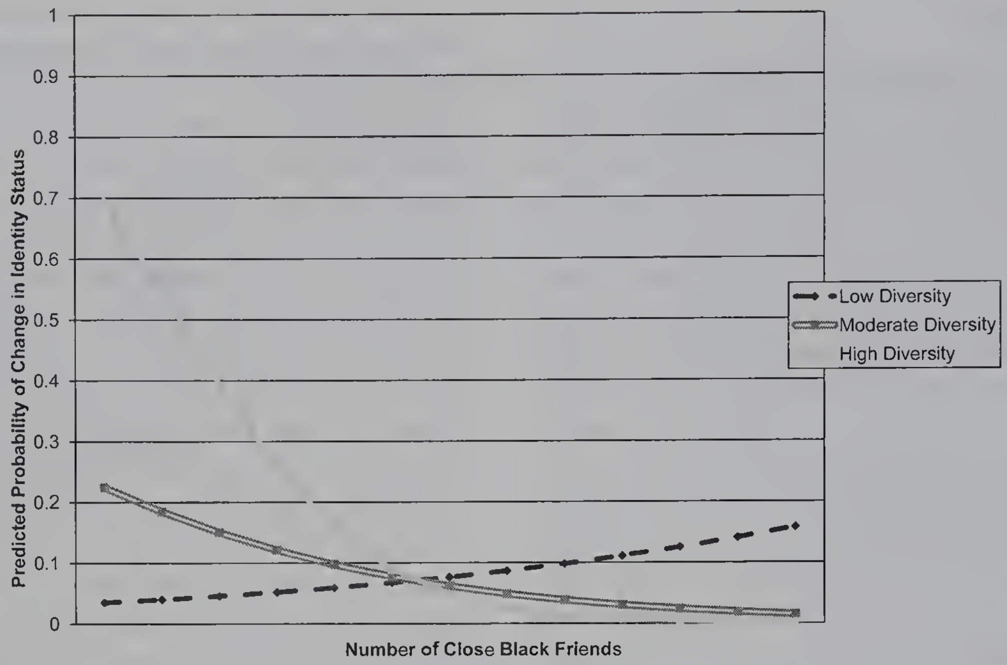


Figure 1. Change from T1 to T2 with T1 predictors: Three-way interaction between close Black friends, contact, and diversity—low contact with Blacks in classes and clubs.

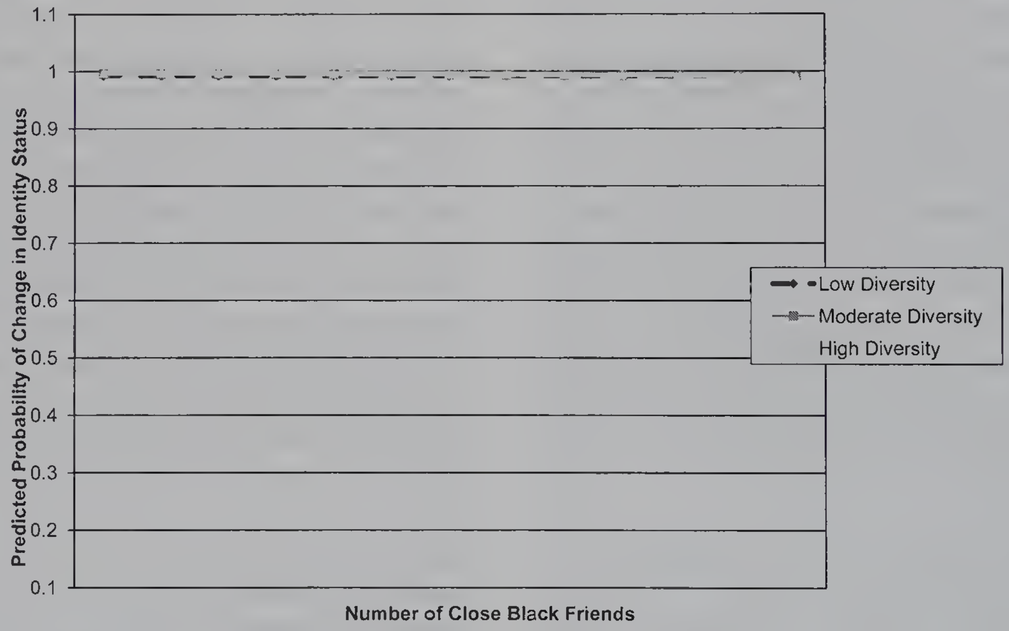


Figure 2. Change from T1 to T2 with T1 predictors: Three-way interaction between close Black friends, contact, and diversity—high contact with Blacks in classes and clubs.

(Figure 2), school diversity and number of close Black friends had no association with the probability of identity change over time.

Examining the same relationships to predict change in racial identity from Time 2 to Time 3, there was evidence of a two-way interaction between school diversity and number of close Black friends (Table 3). As illustrated in Figure 3, as students reported an increase in number of close Black friends, students in low diversity schools (i.e., predominantly White) reported being more likely to report a change in identity status from Time 2 to Time 3. For students attending high diversity schools (i.e., presence of other minorities), however, it seems that reporting more close Black friends was associated with a decreased likelihood of change in identity status.

Analyses were also conducted examining change during the 2-year period between Time 1 and Time 3, and unlike the previous set of analyses, an effect was observed for number of close White friends (Table 4). Namely, there was a main effect of

Table 3
Logistic Regression Predicting Probability of Identity Status Change From Time 2 to Time 3

	B	*SE*
Constant	2.46	2.00
Age	−0.07	0.14
Gender	−0.50	0.32
School diversity (mean Time 1, 2, 3)	0.60	1.72
Close White friend—Time 2	0.01	0.24
Close Black friend—Time 2	−0.21	0.22
Contact—Time 2	0.18	0.17
School Diversity × Close White Friend—Time 2	−4.35	2.98
School Diversity × Close Black Friend—Time 2	−6.17*	2.92
School Diversity × Contact—Time 2	0.28	1.85
School Diversity × Contact × White Friend—Time 2	2.42	3.01
School Diversity × Contact × Black Friend—Time 2	0.14	1.99

*$p < .05$.

Table 4
Logistic Regression Predicting Probability of Identity Status Change From Time 1 to Time 3

	B	*SE*
Constant	3.27	2.06
Age	−0.16	0.14
Gender	−0.32	0.32
School diversity (mean Time 1, 2, 3)	−0.03	1.71
Close White friend—Time 1	1.01**	0.31
Close Black friend—Time 1	0.81**	0.28
Contact—Time 1	0.10	0.18
School Diversity × Close White Friend—Time 1	3.81	3.14
School Diversity × Close Black Friend—Time 1	1.81	2.73
School Diversity × Contact—Time 1	−3.31	2.10
School Diversity × Contact × White Friend—Time 1	−8.24*	3.92
School Diversity × Contact × Black Friend—Time 1	−5.42	3.42

*$p < .05$. **$p < .01$.

number of close White friends where having more White friends was associated with an increased probability of identity change. As well, there was a main effect of having close Black friends associated with an increased likelihood of identity change. Finally, there was also evidence for a three-way interaction between contact with Black students, school diversity, and number of close White friends. To depict this interaction, the sample was divided using a median split of contact with Black students (Figures 4 and 5). As illustrated in Figure 4, students who reported low contact with Black students were less likely to report a change in their identity status if they attended low diversity schools (i.e., predominantly White) and reported having many close White friends. Alternatively, among students reporting high contact with Black students (Figure 5), attending a predominantly White school was associated with a decreased probability of identity change for adolescents reporting having few close White friends.

Discussion

The current study explored how contact with Black students in classes and clubs, close friendships with Black and White students, and school-level

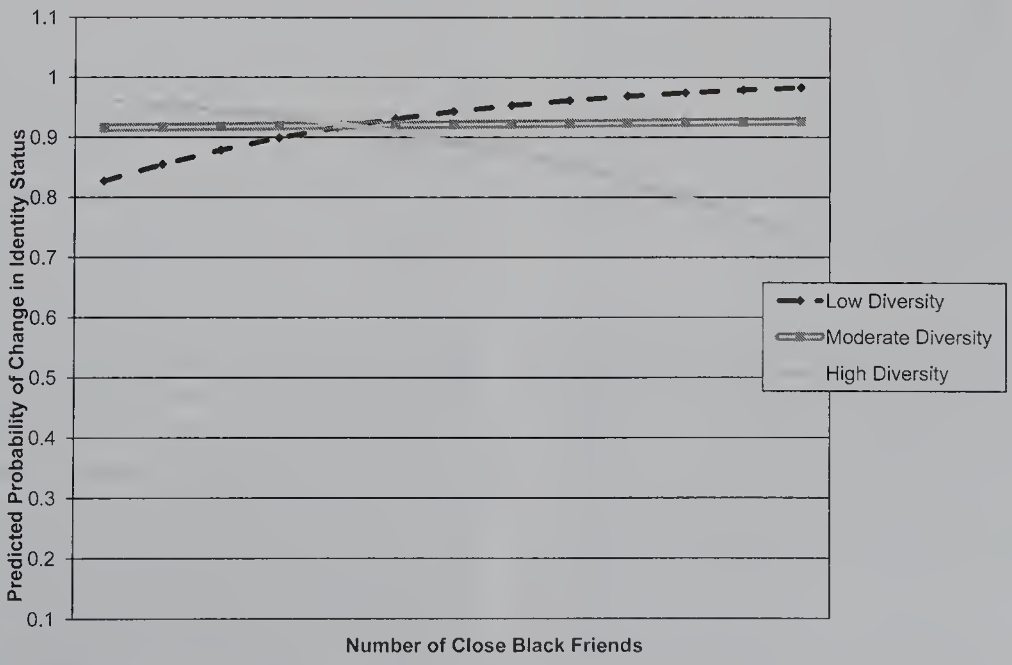


Figure 3. Change from T2 to T3 with T2 predictors: Two-way interaction between diversity and close Black friends.

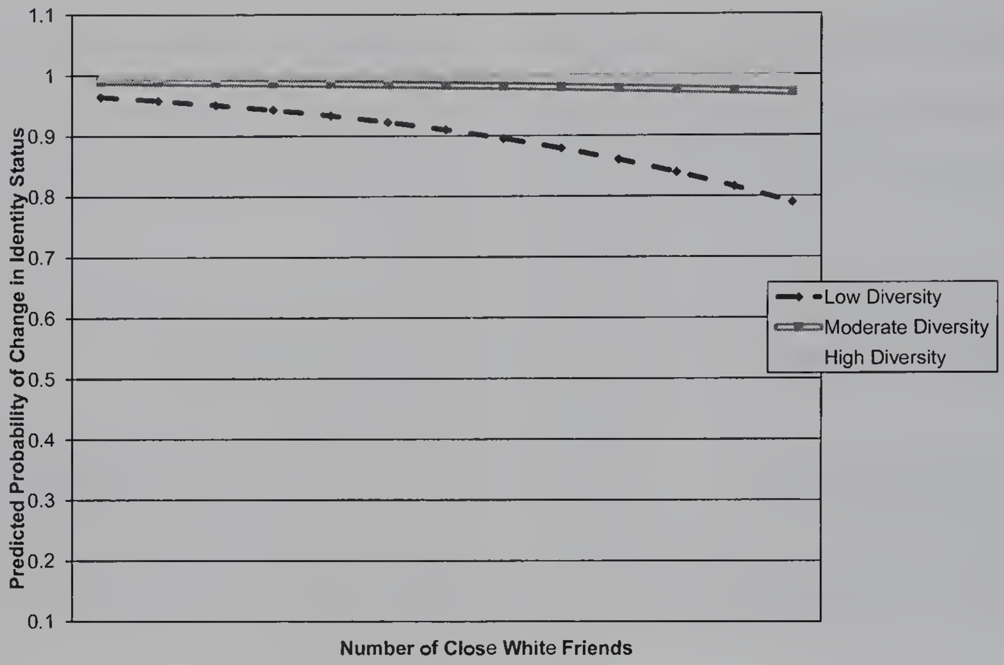


Figure 4. Change from T1 to T3 with T1 predictors: Three-way interaction between close White friends, contact, and diversity—low contact with Blacks in classes and clubs.

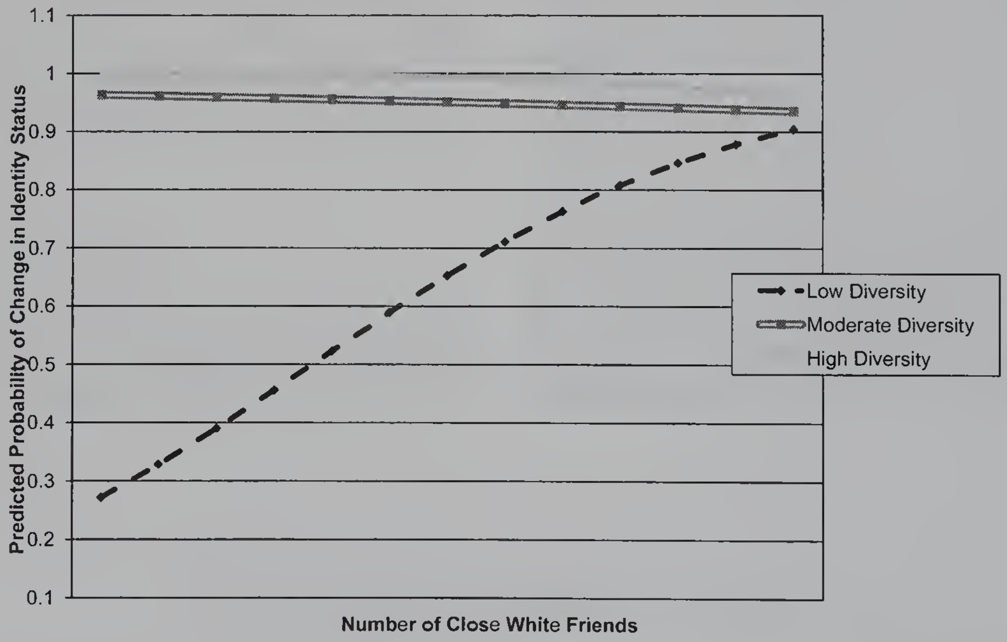


Figure 5. Change from T1 to T3 with T1 predictors: Three-way interaction between close White friends, contact, and diversity—high contact with Blacks in classes and clubs.

diversity independently and jointly predict whether African American adolescents will report change or stability in their racial identity status over 1- and 2-year periods. The data suggested that there were independent effects of contact with Black peers, close White friends, and close Black friends. Interestingly, however, school-level diversity did not appear to have a main effect on the probability of change or stability in racial identity status. Instead, school diversity was observed to have a joint influence with close friendships and contact in classes and clubs. These results point to the active role that youths play in their own identity development. That is, the school setting does not shape racial identity development from a top-down approach; rather, the adolescent creates his or her developmental context through choices like one's close friendship network. To parallel the analyses, the results are discussed in the same order in which they were analyzed: first examining change between adjacent time points (i.e., T1 to T2, T2 to T3), then examining change over a 2-year period (i.e., T1 to T3).

Turning first to the change between T1 and T2, the results indicate a main effect of increased contact with Blacks in classrooms and clubs as linked to an increased likelihood of changing racial identity status among African American youth. This finding is consistent with other research indicating that racial identity was strengthened in the presence of same-race peers. For example, more in-group peer interactions were associated with a stronger sense of ethnic identification among diverse youth (Phinney, Horenczyk, Liebkind, & Vedder, 2001) and same-race peer relationships were associated with resolutions in racial identity formation among African American college students (Tatum, 2004). The present findings and past research suggest that increased contact with same-race peers is linked to changes in racial identity status over time. Although, in the current study, it is unknown if the nature of this change is progression or regression (as outlined in the identity status model; Marcia, 1966; Phinney, 1992), change is occurring under the influence of Black peers suggesting that same-race peers have an important influence on racial identity development.

In addition to this main effect, contact with Blacks in classes and clubs also appeared to interact with school diversity and number of close Black friends to predict change or stability in racial identity between the 1st and 2nd years of the study. Specifically, African American youth with few close Black friends in high diversity settings were more likely to change their racial identity status from Time 1 to Time 2. Recall that low-diversity settings are characterized as predominantly White schools whereas high-diversity settings offer the presence of both same and different race peers. We suggest that this finding indicates a mismatch between the adolescent's choice of friends and what the context allows with respect to opportunities for forming friendships. As suggested by the correlations between close Black and White friends, it seems that youths with few Black friends may have more White friends, but the increased numbers of Blacks in classes and clubs affords new opportunities for racial encounters. Nigresence theory posits that racial encounters result in individuals reevaluating the role that race plays in their lives, which results in an advanced and nuanced racial identity for African Americans (Cross & Fhagen-Smith, 2001). Nigresence theory suggests that racial encounters may be negative (e.g., racially discriminatory experiences) or positive (e.g., making the soccer team); therefore, youths with few Black friends may be experiencing positive encounters as they come into increased contact with their African American peers. These encounters may be more probable in high-diversity settings given the increased numbers of other African American youth, and they may result in increased exploration or commitment, which would be coded as a change in racial identity status in the current study.

Turning to the analyses examining predictors of the change from T2 to T3, the data indicate interactive effects between school diversity and having more close Black friends. Specifically, in high-diversity settings, youths who reported having many close Black friends also reported that their racial identity status was less likely to change. The findings suggest that intraracial friendships have differential influences on racial identity development depending on the level of diversity in the setting. One explanation for this finding is that low-diversity settings increase the likelihood of negative interactions (i.e., discriminatory treatment due to minority status), which may be linked to racial identity changes. Previous research indicates that greater school diversity was related to less positive perceptions of the general school climate (Benner, Graham, & Mistry, 2008) given that discrimination is believed to be maximized in evenly mixed racial contexts (Sigelman, Bledsoe, Welch, & Combs, 1996). Additional research indicates that perceptions of racially discriminatory acts by peers were associated with increased identity exploration levels over time among African American youth (Pahl & Way, 2006). Thus, low diversity levels may be associated with more perceptions of discrimination among African American youth, and these perceptions of racial discrimination may be influencing the increased likelihood of changing racial identity status groups. A second explanation is the quality of contact that African American youth have with Black and White friends. Diverse settings afford Black youth more opportunities to have White friends, but previous research suggests that ethnic minorities had less positive interactions with White friends than with friends of their same racial background (Shelton & Richeson, 2006). Yet, African American children who had more contact with Black friends considered the contact to be positive in that it helped prepare them for discriminatory treatment (Rowley et al., 2008). Thus, the nature of the interactions is distinct depending on the race of the friends and research suggests that less positive contact increases with more White friends. As such, these interactions may influence the degree to which African American youth change their racial identity status groups.

Finally, examining change over a 2-year period from T1 to T3, the data suggested an interaction between context, contact with Black students, and close friendships with White students. It appears that being in a predominantly White setting is an important factor for racial identity change when African American youths report having a lot of White friends but little contact with Blacks in their classes and clubs. Similarly, a predominantly White setting matters when youths have few close White friends and a lot of contact with Blacks in their classes and clubs. In both groups, youths are in predominantly White settings, but one group associates with a lot of White friends whereas the other group does not, and there is a decreased likelihood of racial identity change in both scenarios. One explanation is related to a tenet of the contact hypothesis, which posits that more intergroup contact results in more attention toward out-group members (Pettigrew, 1998). The nature of the predominantly White setting may result in increased awareness and attention to interactions with White out-group as opposed to Black in-group members. Hence, this focus may decrease the probability of change in the racial identity status groups among African American youth, as exploration and commitment related to being African American are minimized. The present findings are contrary to previous research conducted with White (Hamm, 2000) and Latino youths (Umana-Taylor, 2004), showing that cross-race friendships were influential for ethnic identity. In the present sample, students reported that about "half" of their friends were of each group on both the close Black friend and close White friend variable. Therefore, it seems as though most of the students in this sample did have both in-group and out-group friendships regardless of the diversity of the school. Despite this similarity, it seems that only close friendships with same-race peers interacted with school diversity to influence changes in youths' racial identity over time. This may be true because this study was conducted in a predominantly White town, so the significance and meaning of close Black friends was more influential for changes in racial identity, a finding consistent with prior research (Phinney, Romero, et al., 2001; Tatum, 2004).

Although the current study offers a more nuanced and detailed examination of the relations between contact, context, and identity change, there are a few limitations that require consideration. Although cluster analysis provides a valuable method for classifying individuals, the technique is sample dependent. An additional limitation concerns the lack of assessment for previous development in that cluster analysis does not account for previous exploration or commitment levels. Another consideration is that the current study only examined how school-based peer interactions and diversity influence racial identity outcomes and research suggests that only 28% of youths who report out-group friends in school also report having out-group friends outside of school (DuBois & Hirsch, 1990). Therefore, it would seem important for future research to consider interracial and interracial interactions that occur outside of the school context. In addition, the time period for change is unclear given that the data were collected once a year. Despite the fact that the observed changes in identity status at 1- and 2-year intervals are consistent with developmental theories suggesting adolescence is an important time for identity exploration and construction (Erikson, 1968), data collected at smaller intervals might suggest more stability or change. Similarly, data collected during smaller time intervals might provide more information about how antecedents (i.e., school diversity) are related to subsequent changes in identity status. It should also be noted that the sample size was too small to allow for more detailed analyses to identify subgroups of identity change trajectories (e.g., regressing from achieved to foreclosed vs. regressing from achieved to moratorium). For example, it is plausible that the interaction of school diversity and Black friends may be associated with specific trajectories. It is also important for future research to consider the content and quality of interracial and intraracial interactions and how these dimensions contribute to youths' thoughts and feelings about their racial group membership. Finally, the current study only examines the racial dimension of adolescents' identity development and does not capture the possibility that although adolescents may not be changing in their levels of commitment and exploration of their *racial* identity, they could be changing in terms of developing a bicultural or multicultural perspective.

Nevertheless, the current study represents an important and unique contribution to the racial identity development literature by assessing intergroup and intragroup contact and school diversity as antecedents of change. Although the contact hypothesis posits that contact will have a positive impact on reducing prejudice and bias toward out-group members (Pettigrew & Tropp, 2006), the results suggest that intergroup contact has important implications for racial identity development among African American youths. Specifically, school diversity and having Black friends influence whether African American youths change racial identity status groups or remain stable across time. As racial identity has been shown to be an important construct for African American and minority youths, this study advances our understanding of key antecedents for this integral developmental process.

References

Alba, R., & Nee, V. (2003). *Rethinking the American mainstream: Assimilation and contemporary immigration*. Cambridge, MA: Harvard University Press.

Benner, A. D., Graham, S., & Mistry, R. S. (2008). Discerning direct and mediated effects of ecological structures and processes on adolescents' educational outcomes. *Developmental Psychology*, *44*, 840–854.

Bracey, J. R., Bamaca, M. Y., & Umana-Taylor, A. J. (2004). Examining ethnic identity and self-esteem among biracial and monoracial adolescents. *Journal of Youth and Adolescence*, *33*, 123–132.

Cross, W. E., Jr., & Fhagen-Smith, P. (2001). Patterns of African American identity development: A life span perspective. In B. Jackson & C. Wijeyesinghe (Eds.), *New perspectives on racial identity development: A theoretical and practical anthology* (pp. 243–270). New York: New York University Press.

de Souza Briggs, X. (2007). "Some of my best friends are …": Interracial friendships, class, and segregation in America. *City & Community*, *6*, 263–290.

DuBois, D. L., & Hirsch, B. J. (1990). School and neighborhood friendship patterns of Blacks and Whites in early adolescence. *Child Development*, *61*, 524–536.

Dutton, S. E., Singer, J. A., & Devlin, A. S. (1998). Racial identity of children in integrated, predominantly White, and Black schools. *Journal of Social Psychology*, *138*, 41–53.

Erikson, E. (1968). *Identity: Youth and crisis*. New York: Norton.

Hair, J. F. J., & Black, W. C. (2000). Cluster analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 147–205). Washington, DC: American Psychological Association.

Hamm, J. V. (2000). Do birds of a feather flock together? The variable bases for African American, Asian American, and European American adolescents' selection of similar friends. *Developmental Psychology*, *36*, 209–219.

Juvonen, J., Nishina, A., & Graham, S. (2006). Ethnic diversity and perceptions of safety in urban middle schools. *Psychological Science*, *17*, 393–400.

Kao, G., & Joyner, K. (2006). Do Hispanic and Asian adolescents practice panethnicity in friendship choices? *Social Science Quarterly*, *87*, 972–992.

Kao, G., & Vaquera, E. (2006). The salience of racial and ethnic identification in friendship choices among Hispanic adolescents. *Hispanic Journal of Behavioral Sciences*, *28*, 23–47.

Kubitschek, W. N., & Hallinan, M. T. (1998). Tracking and students' friendships. *Social Psychology Quarterly*, *61*, 1–15.

Lee, R. M., Noh, C.-Y., Yoo, H. C., & Doh, H.-S. (2007). The psychology of diaspora experiences: Intergroup contact, perceived discrimination, and the ethnic identity of Koreans in China. *Cultural Diversity and Ethnic Minority Psychology*, *13*, 115–124.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, *3*, 551–558.

Marcia, J. E. (1980). Identity in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 159–187). New York: Wiley.

Masson, C. N., & Verkuyten, M. (1993). Prejudice, ethnic identity, contact and ethnic group preferences among Dutch young adolescents. *Journal of Applied Social Psychology*, *23*, 156–168.

McMahon, S. D., & Watts, R. J. (2002). Ethnic identity in urban African American youth: Exploring links with self-worth, aggression, and other psychosocial variables. *Journal of Community Psychology*, *30*, 411–431.

Mickelson, R. A., & Heath, D. (1999). The effects of segregation and tracking on African American high school seniors' academic achievement. *Journal of Negro Education*, *68*, 566–586.

Molina, L. E., & Wittig, M. A. (2006). Relative importance of contact conditions in explaining prejudice reduction in a classroom context: Separate and equal? *Journal of Social Issues*, *62*, 489–509.

Nier, J. A., Gaertner, S. L., Dovidio, J. F., Banker, B. S., Ward, C. M., & Rust, M. C. (2001). Changing interracial evaluations and behavior: The effects of a common group identity. *Group Processes & Intergroup Relations*, *4*, 299–316.

Pahl, K., & Way, N. (2006). Longitudinal trajectories of ethnic identity among urban Black and Latino adolescents. *Child Development*, *77*, 1403–1415.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Pettigrew, T. F. (2006). The advantages of multilevel approaches. *Journal of Social Issues*, *62*, 615–620.

Pettigrew, T. F. (2007). Future directions for intergroup contact theory and research. *International Journal of Intercultural Relations*, *32*, 187–199.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

Phinney, J. S. (1989). Stages of ethnic identity in minority group adolescents. *Journal of Early Adolescence*, *9*, 34–49.

Phinney, J. S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research, 7*, 156–176.

Phinney, J. S. (1993). A three-stage model of ethnic identity development in adolescence. In M. E. Bernal (Ed.), *Ethnic identity: Formation and transmission among Hispanics and other minorities. SUNY series, United States Hispanic studies* (pp. 61–79). Albany: State University of New York Press.

Phinney, J. S., Horenczyk, G., Liebkind, K., & Vedder, P. (2001). Ethnic identity, immigration, and well-being: An interactional perspective. *Journal of Social Issues, 57*, 493–510.

Phinney, J. S., Romero, I., Nava, M., & Huang, D. (2001). The role of language, parents, and peers in ethnic identity among adolescents in immigrant families. *Journal of Youth and Adolescence, 30*, 135–153.

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press.

Quillian, L., & Campbell, M. E. (2003). Beyond black and white: The present and future of multiracial friendship segregation. *American Sociological Review, 68*, 540–566.

Roberts, R. E., Phinney, J. S., Masse, L. C., Chen, Y. R., Roberts, C. R., & Romero, A. (1999). The structure of ethnic identity of young adolescents from diverse ethnocultural groups. *Journal of Early Adolescence, 19*, 301–322.

Rowley, S. J., Burchinal, M. R., Roberts, J. E., & Zeisel, S. A. (2008). Racial identity, social context, and race-related social cognition in African Americans during middle childhood. *Developmental Psychology, 44*, 1537–1546.

Schwarzwald, J., & Cohen, S. (1982). Relationship between academic tracking and the degree of interethnic acceptance. *Journal of Educational Psychology, 74*, 588–597.

Seaton, E. K., Scottham, K. M., & Sellers, R. M. (2006). The Status model of racial identity development in African American adolescents: Evidence of structure, trajectories and well-being. *Child Development, 77*, 1416–1426.

Seaton, E. K., Yip, T., & Sellers, R. M. (2009). A longitudinal examination of racial identity and racial discrimination among African American adolescents. *Child Development, 80*, 406–417.

Shelton, J. N., & Richeson, J. A. (2006). Ethnic minorities' racial attitudes and contact experiences with White people. *Cultural Diversity and Ethnic Minority Psychology, 12*, 149–164.

Sigelman, L., Bledsoe, T., Welch, S., & Combs, M. W. (1996). Making contact? Black-White social interaction in an urban setting. *American Journal of Sociology, 101*, 1306–1332.

Tatum, B. D. (1997). *"Why are all the Black kids sitting together in the cafeteria?" and other conversations about race*. New York: Basic Books.

Tatum, B. D. (2004). Family life and school experience: Factors in the racial identity development of Black youth in White communities. *Journal of Social Issues, 60*, 117–135.

Tropp, L. R. (2007). Perceived discrimination and interracial contact: Predicting interracial closeness among Black and White Americans. *Social Psychology Quarterly, 70*, 70–81.

Umana-Taylor, A. J. (2004). Ethnic identity and self-esteem: Examining the role of social context. *Journal of Adolescence, 27*, 139–146.

Yip, T., Seaton, E. K., & Sellers, R. M. (2006). African American racial identity across the lifespan: Identity status, identity context and depressive symptoms. *Child Development, 77*, 1503–1516.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1445–1459

Do Adolescents and Parents Reconstruct Memories About Their Conflict as a Function of Adolescent Attachment?

Matthew J. Dykas
State University of New York at Oswego

Susan S. Woodhouse
Pennsylvania State University

Katherine B. Ehrlich and Jude Cassidy
University of Maryland

This study examined whether 17-year-old adolescents (n = 189) and their parents reconstructed their memory for an adolescent–parent laboratory conflict over a 6-week period as a function of adolescent attachment organization. It also compared participants' perceptions of conflict over time to observational ratings of the conflict to further characterize the nature of the attachment-related memory biases that emerged. Secure adolescents reconstructed interactions with each parent more favorably over time, whereas insecure adolescents showed less favorable reconstructive memory. Likewise, mothers of secure girls reconstructed conflicts more favorably over time, whereas mothers of insecure boys showed less favorable reconstructive memory. Participant ratings were associated with observational ratings in theoretically consistent ways. Contrary to expectations, fathers showed no attachment-related memory biases.

Developmental researchers have shown considerable interest in understanding the nature of adolescent–parent conflict. The causes of adolescent–parent conflict, for example, have been relatively well studied, and data indicate that such conflict stems from both normative changes within the child–parent relationship (e.g., puberty-related biological changes, increasing adolescent autonomy, and adolescents' reduced social and emotional dependence on parents; Collins & Laursen, 1992; Conger & Ge, 1999; Ge, Conger, & Elder, 1996) as well as dispositional and parenting factors (e.g., temperament, maternal conflict reactions; see Eisenberg et al., 2008). The nature and extent of adolescent–parent conflict varies within families, and although conflict can emerge over issues of "critical" importance and/or urgency, disagreements over everyday "trivial" matters (e.g., chores, curfew) are often the root of adolescent–parent conflict (Smetana, 1996). Interestingly, although conflict often places stress on adolescents and their parents, researchers believe that conflict is part of an adaptive socialization process that promotes adolescent and parent development (Adams & Laursen, 2007; Collins & Laursen, 1992; Smetana, Campione-Barr, & Metzger, 2006).

Although a wealth of evidence indicates that adolescents' conflict with parents is common and developmentally significant, the social-cognitive *aftereffects* of conflict are not well understood. For example, it is reasonable to believe that adolescents and parents remember specific conflict incidents over time, but is their memory for conflict an accurate representation of how they initially perceived the conflict, and how does it mesh with observer ratings of the conflict? A large and convergent body

---

This research was supported by Grant HD36635 from the National Institute for Child Health and Human Development to Jude Cassidy, and by an NRSA Predoctoral Award to Katherine Ehrlich (F31 DA027365-01). Portions of this research were presented at the International Family Violence and Child Victimization Research Conference, July 2006, Portsmouth, New Hampshire. We thank the families who participated in this research and Mindy Rodenberg Cabrera for supervising data collection. We are grateful to Kristen Intlekofer, Daniel Isenberg, Melissa Klein, Laura Long, Fatima Ramos Marcuse, Amber Wong, and Yair Ziv for coding the interactions; and to Inbal Kivenson Bar-On, Mindy Rodenberg Cabrera, June Sroufe, Sue Watson, and Marina Zelenko for coding the interviews used in this study. We are also grateful to Kevin O'Grady and Tracy Tomlinson of the University of Maryland Design and Statistical Analysis Laboratory, and Jacqueline Reihman for their statistical consultation, and to Phillip Shaver and an anonymous reviewer for their helpful comments on earlier drafts of this report.

Correspondence concerning this article should be addressed to Matthew J. Dykas, Department of Psychology, State University of New York at Oswego, Oswego, NY 13126. Electronic mail may be sent to matt.dykas@oswego.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0010

of literature indicates that memories may not always be accurate depictions of interpersonal experience (Baldwin, 1995; Fiske & Taylor, 1991). Thus, over time, when the memory for a conflict event degrades (i.e., as individuals forget the details of these specific interactions), how do adolescents and parents ''fill in the gaps'' or *reconstruct* their memory for the details of the conflict? Grotevant (1998), in his discussion of adolescent–parent relationships, has suggested that it is important to consider ''the ways in which individuals frame events and reconstruct their own pasts . . . even if these reconstructions are not completely faithful to history'' (p. 1126). Moreover, this question about reconstructive memory for conflict has important implications for how adolescent–parent conflict is both understood and treated by researchers and practitioners.

Using Bowlby's (1969/1982, 1973) attachment theory as a conceptual framework, Feeney and Cassidy (2003) began to address novel questions about reconstructive memory for adolescent–parent conflict. They proposed that the way in which adolescents reconstruct their memory for adolescent–parent conflict is governed in part by their internal working models of attachment, which are internalized mental representations of attachment experiences forged through repeated daily secure base interactions with parents and other caregivers (Bowlby, 1973). If these models provide rules for the direction of memory for attachment-related information (as Bowlby, 1973, and Main, Kaplan, & Cassidy, 1985, claimed), then adolescents might tap into these models to reconstruct the degraded memory of their conflicts with parents. More precisely, if these internal working models are secure, adolescents should be inclined to reconstruct their memory for adolescent–parent conflict such that they remember these interactions as more positive than they did in their original perceptions (i.e., in a more positive and less negative manner), because they are relying on secure internalized attachment representations. In contrast, if these models are insecure, adolescents should reconstruct their memory for conflictual interactions, such that they will remember these interactions as more negative than their original perceptions (e.g., in a more negative and less positive manner) because they are relying on insecure attachment representations. These propositions are consistent with other theories suggesting that individuals recall events from memory in ways that reflect their views and perceptions (see Pasupathi, 2001; Tversky & Marsh, 2000), and broader theories that individuals' mental representations mediate longitudinal connections between early experience and later functioning (see Dweck & London, 2004).

The first empirical examination of these ideas did not focus on individuals' attachment organization (described next) but instead involved the use of adolescents' self-reported attachment-related representations of their mothers and fathers. Feeney and Cassidy's (2003) results from both their initial study and a replication study revealed that, as expected, adolescents reconstructed their memory of adolescent–parent conflict over time as a function of their representations of a particular parent. For example, in relation to separate laboratory-based conflict discussions with mothers and fathers, when adolescents' perceptions of the parent were more negative (e.g., as adolescents responded to questionnaires by describing their mothers or fathers as unavailable for help when needed or as not understanding their feelings), adolescents were more likely to remember the discussion with that particular parent as *less positive* and *more negative* than reported 6 weeks earlier, just after the conflict episode.

In light of this intriguing evidence that adolescents reconstruct their memory for adolescent–parent conflict as a function of their representations of parents, the purpose of the present investigation was to use Feeney and Cassidy's (2003) sample to extend this novel line of research in three noteworthy ways. First, Feeney and Cassidy examined adolescents' self-reported representations of each parent in relation to a conflict interaction with that particular parent (e.g., asking whether an adolescent's representation of mother is linked to the adolescent's memory for a conflict task with her). In the present investigation, instead of examining adolescents' representations of each specific parent, we examined whether and how adolescents' Adult Attachment Interview (AAI; George, Kaplan, & Main, 1996) classifications, which were not available at the time of the Feeney and Cassidy report, were linked to adolescents' memory for conflict with parents. Our use of the AAI in this investigation is both conceptually and methodologically distinct from Feeney and Cassidy's earlier work because the AAI taps an adolescent's attachment organization (referred to as the individual's ''current state of mind with respect to attachment''; Main & Goldwyn, 1998) rather than representations of individual parents. This ''current state of mind'' is thought to be a relatively stable global attachment organization that emerges in adolescence based on the consolidation and integration of internalized experiences with mother, father, and other attachment figures (Allen, 2008; Allen, Boykin-McElhaney, Kuperminc, & Jodl, 2004; Bretherton & Munholland, 1999, 2008; Hesse, 1999, 2008). Thus, from a theoretical perspective, this study permits examination of one of the core propositions of attachment theory (Bowlby, 1973): the idea that a link exists between attachment and the processing of social information. This proposition is supported by considerable empirical evidence from studies of both children and adults (e.g., Belsky, Spritz, & Crnic, 1996; Kirsh & Cassidy, 1997; van Emmichoven, van IJzendoorn, de Ruiter, & Brosschot, 2003). Yet, despite the fact that this proposition is so theoretically central, researchers have examined it surprisingly little, especially during adolescence (see Dykas & Cassidy, 2007, for a review). Moreover, from a methodological perspective, the findings related to the AAI can be compared and contrasted to the many other AAI-related findings in the literature.

Second, we extend the Feeney and Cassidy (2003) study by examining not only how attachment is linked to *adolescents'* reconstructive memory for conflict, but also to *mothers'* and *fathers' reconstructive memory*. Researchers have shown considerable interest in understanding whether and how parental cognitions influence child functioning (e.g., Bugental & Happaney, 2002) and a central component of attachment theory is the notion that the way in which parents process attachment-related information contributes to their children's quality of attachment to them, mediated largely by parenting behavior (Bowlby, 1988; van IJzendoorn, 1995; see also George & Solomon, 2008). It is thought, for example, that parents who process information about their children in a negatively biased way will have difficulty serving as a secure base and/or safe haven for their children. This lack of a parental secure base and/or safe haven, in turn, is thought to contribute to an insecure attachment. There is considerable evidence that parents' representations of their children are related to their parenting behavior and/or their children's attachment to them (Benoit, Parker, & Zeanah, 1997; Grienenberger, Kelly, & Slade, 2005; Slade, Belsky, Aber, & Phelps, 1999; see George & Solomon, 2008, for a review), yet to our knowledge no published study has examined links between adolescent attachment and parents' reconstructive memory, or social information processing more generally. Importantly, we note that simply the inclusion of fathers in the present investigation is in itself an advance over nearly all adolescent attachment research, which has largely focused on mothers (see Grotevant, 1998, for discussion of the importance of including fathers in developmental research).

Finally, we extend Feeney and Cassidy's (2003) previous work by investigating how memory for adolescent–parent conflict relates to the *observed behavioral quality* of the conflict as a function of adolescent attachment organization. More precisely, using observational data collected during the adolescent–parent conflict interactions, which were not available at the time of the Feeney and Cassidy report, we were able to discern whether insecure and secure adolescents' (as well as their parents') memory for conflict became more or less attuned over time to observer ratings of behaviors exhibited in the conflict. This final research extension is of considerable theoretical interest because it lends additional insight into whether memory for adolescent–parent conflict, over time, converges or diverges from the observed behavior during the conflict interaction as a function of adolescents' attachment organization. For example, it is possible that a secure or an insecure individual could have an overly favorable or unfavorable initial memory of an event in comparison to observers that ultimately becomes more attuned to the observed quality of the event over some time period. On the other hand, it is possible that an individual's view of an event could initially match observers' perspectives on the event, but later memory could diverge from the observers' perspective.

On the basis of both attachment theory and Feeney and Cassidy's (2003) earlier findings, we examined the following two hypotheses. First, we hypothesized that the memory processes of secure and insecure adolescents differ, such that secure adolescents are more likely to have a bias to remember adolescent–parent conflictual interactions more positively and less negatively over time; in contrast, insecure adolescents were hypothesized to have a bias to remember the conflict interaction more negatively and less positively over time. Second, we hypothesized that the memory processes of parents of secure and insecure adolescents differ, such that parents of secure adolescents are more likely to have a bias to remember the adolescent–parent conflictual interactions less negatively and more positively over time; in contrast, parents of insecure adolescents were hypothesized to show a bias to remember the conflict interaction more negatively and less positively over time. For each hypothesis, we conducted accompanying examinations to investigate an additional research question about whether insecure and secure adolescents' (and their parents') memory for conflict becomes less or more attuned over time to the coder-rated behaviors exhibited in the conflict. In each analysis, we included adolescent gender as an additional factor, given an interest in studying unique characteristics of specific parent–adolescent dyads (e.g., mother–daughter, father–son; see Russell & Saebel, 1997).

Method

Participants

Participants were 189 eleventh-grade students recruited from seven public suburban high school and their parents. (We combined the cohorts in Feeney & Cassidy, 2003, Study 1 and Study 2 to obtain greater statistical power; see Table 1 for sample demographic information.) Data reported in this investigation were gathered from a larger study of adolescents' social relationships. Because of a focus on adolescent relationships with both mothers and fathers, we recruited English-speaking participants with married parents. Adolescents were paid $125 for participating in the larger study. Sample size varies across analyses due to missing data.

Procedure

We gathered data during three sessions spanning approximately 6 weeks. First, adolescents visited our laboratory and engaged, separately with their mother and father, in 10-min conflict discussions. Adolescents and their parents used a checklist to rate how much they disagreed with each other about 19 topics that parents and teens frequently disagree about, such as "chores," "homework," and "talking back to parents." A research assistant chose three topics for the mother–adolescent discussion and three topics for the father–adolescent discussion, selecting topics that were rated by the parent and adolescent as high in disagreement (see Feeney & Cassidy, 2003, for a further description of the generation of adolescent–parent conflict discussions). The order in which adolescents participated in the discussions with each parent was counterbalanced, and all discussions were videotaped for later coding. After each discussion, adolescents and parents completed the Emotional Response to Conflict Scale (ERCS; Cassidy, 1998; T1); adolescents also completed the vocabulary subtest of the Shipley Institute of Living Scale (Shipley, 1946) during this first session. In the second session (1 month later), adolescents completed the AAI in a different university laboratory room. Finally, 6 weeks after the adolescent–parent conflict discussions, we mailed a follow-up ERCS (T2) to adolescents and their parents. Participants were asked to complete questionnaires independently and to return them via mail in the envelope provided.

Table 1
Sample Characteristics

	Total ($n = 189$)
Adolescent gender (%)	
Male	38
Female	62
Family race (%)	
White/Caucasian	73
Black/African American	14
Asian	10
Hispanic	3
Annual household income (%)	
–$60,000	15
$61,000+	85
Mother education (%)	
High school	23
College/graduate degree	77
Father education (%)	
High school	18
College/graduate degree	82
Adolescent Shipley scores	$M = 29, SD = 4.8$

Measures

Emotional Response to Conflict Scale (Cassidy, 1998). This 31-item scale was designed to assess individuals' emotional responses to a 10-min laboratory conflict discussion task (Strodtbeck, 1951), both immediately afterward in the laboratory, and 6 weeks later at home. The items contained in this questionnaire fell into eight indices that, following Feeney and Cassidy (2003), we subsequently reduced into three summary scores (see Feeney & Cassidy, 2003, for a full description of the ERCS and for the theoretical and factor analytic basis of this data-reduction strategy). First, we created a *positive interaction* score by averaging the items in the positive discussion, own positive emotions, other positive emotions, and positive treatment indices (mean $\alpha = .87$). Second, we created a *negative interaction* score by averaging the items in the negative discussion, own negative emotions, and other negative emotions indices (mean $\alpha = .84$). Finally, we created a *hostile treatment* score by averaging the hostile treatment item ratings (mean $\alpha = .76$). For the entire sample across both time points, associations between these three summary scores ranged from $r = -.42$ (positive interaction and hostile treatment scores) to $r = .66$ (negative interaction and hostile treatment scores).

Adolescent–Parent Conflict Interaction Coding System (Ziv, Cassidy, & Ramos-Marcuse, 2002). We used this coding system, based in part on an earlier system developed by Kobak, Cole, Ferenz-Gillies, Fleming, and Gamble (1993), to assess adolescent and parent verbal and nonverbal (e.g., facial) behavior during a 10-min laboratory discussion task. In this study, we analyzed scales assessing adolescent *secure base use* (i.e., the degree to which the adolescent maintained "secure relatedness" during the discussion through the demonstration of relationship-maintaining behaviors even when under the stress of discussing a disagreement) and parents' secure base provision (i.e., the degree to which the parent was sensitive, responsive, and available to the adolescent throughout the conflict discussion). We also analyzed separate scales for adolescents and their parents that assessed each individual's *avoidance of discussing disagreement* (i.e., the overall degree to which the individual actively disengaged from the discussion of conflict) and *hostility* (i.e., the overall degree to which the individual engaged in sarcastic/contemptuous comments or smiles, dysfunctional anger, disgust, and/or aggressive posturing). Finally, we analyzed the *open communication* dyadic scale that assessed the overall degree to which the adolescent–parent dyad demonstrated comfort and openness in discussing their thoughts, feelings, and emotions during the conflict in a mutually cohesive and fluid manner. All of these scales were based on a 7-point rating system (with appropriate behavioral anchors); six highly trained coders, blind to additional data regarding the participants, independently rated videotaped discussions. Agreement was assessed continuously throughout the coding period and at least two coders coded a randomly selected 17% of adolescent–mother interactions ($n = 32$) and 16% of father–adolescent interactions ($n = 31$). Sufficient reliability emerged for this group of coders: Intraclass correlation coefficients ranged from .76 on mothers' secure base provision to .84 on adolescents' hostility toward father. Disagreements were conferenced and consensus scores were used in all analyses.

Because high correlations emerged within and between the adolescent and parent behavior scale score sets (*r*s ranging from −.22 to .81), we conducted principal component analyses (PCAs) to combine these data. (We conducted separate PCAs for behaviors rated in the adolescent–mother conflict and for the behaviors rated in the adolescent–father conflict. We did not include participants' behavioral hostility scores in these PCAs because they corresponded directly to the hostile treatment scores derived from the ERCS.) In the PCA for adolescent–mother conflict, all behaviors loaded onto a single factor explaining 68.7% of the variance (eigenvalue = 2.75). Similarly, in the PCA for adolescent–father conflict, all behaviors loaded onto a single factor explaining 64.1% of the variance (eigenvalue = 2.56). On the basis of these two PCAs and the positively scaled factor loadings, we computed standardized factor scores based on the factor loadings, which we labeled *positive adolescent–mother dyadic conflict behavior* scores and *positive adolescent–father dyadic conflict behavior* scores.

Adult Attachment Interview (George et al., 1996). We used this well-known semistructured interview to assess adolescents' "current state of mind with respect to attachment," focusing principally on memories of attachment-related experiences during childhood (see Hesse, 1999, 2008, for full descriptions of the AAI). Interviews lasted approximately 1 hr and were audiotaped for later verbatim transcription. We made minor modifications to this interview to make some of the questions more appropriate for an adolescent population (Allen, Moore, Kuperminc, & Bell, 1998). The psychometric properties of the AAI have been well established (see Hesse, 1999, 2008). Although the AAI was originally developed to assess adult attachment, a wealth of research indicates that the AAI is a valid measure of adolescents' attachment representations as well (see Allen, 2008, for a review). Moreover, the AAI has shown considerable test–retest reliability in several short-term and long-term longitudinal studies (see Hesse, 1999, 2008), and Allen et al. (2004) reported that AAI attachment security is stable across adolescence. Following Allen (2008), we assumed that all adolescents in this study possessed a "current state of mind with respect to attachment" that could be assessed using the AAI.

Using Main and Goldwyn's (1998) classification system, coders rated each transcript on a series of 9-point scales. Coders used inferred experience scales to rate the likely quality of adolescents' childhood attachment-related experiences, and "state of mind" scales (e.g., coherence of mind) to assess adolescents' current state of mind with respect to attachment. Based on an integrated consideration of both the adolescents' inferred experiences and state of mind, coders assigned one of four principal classifications to the transcript. Adolescents classified as secure/autonomous coherently described various childhood experiences, valued attachment relationships, and considered attachment-related experiences as important to personal development. Adolescents classified in any of the three insecure groups demonstrated an inability to describe their childhood attachment-related experiences coherently. Specifically, adolescents classified as insecure/dismissing were derogatory about the importance of attachment and/or dismissed the impact that early negative experiences had on personal development and on attachment relationships. Adolescents classified as insecure/preoccupied demonstrated an excessive, confused/passive, and nonobjective preoccupation with attachment relationships and/or experiences. Adolescents classified as insecure/unresolved revealed lapses in the monitoring of reasoning or discourse when discussing loss or trauma. When transcripts could not be placed into any other category, they were labeled ''insecure/cannot classify.''

Four coders who were trained and certified as reliable by Mary Main and Erik Hesse (developers of and/or experts on the AAI and classification system) coded the AAI transcripts. All four coders were blind to additional information regarding participants. Interrater reliability among these coders was assessed continuously throughout coding; a randomly selected 29% of cases ($n = 55$) were coded by at least two coders (78% agreement, $\kappa = .61$, $p < .05$). All disagreements were resolved by a fifth independent coder who coded no additional data.

Shipley Institute of Living Scale (Shipley, 1946). This 40-item vocabulary subtest, in which adolescents were asked to select the correct synonym from a list of four possibilities, was used to assess adolescents' verbal knowledge (to determine whether it was related to adolescents' AAI group classifications). The psychometric properties of this measure are well established (Kirk & Rattan, 1992).

Results

Descriptive Statistics

AAI classifications. The distribution of the adolescent AAI classifications was: 126 secure/autonomous (67%), 44 insecure/dismissing (23%), 10 insecure/preoccupied (5%), 6 unresolved (3%), and 2 insecure/cannot classify (1%). (Scheduling difficulties prohibited 1 adolescent from completing the AAI.) This distribution is similar to distributions obtained in other studies of low-risk adolescents (e.g., Ammaniti, van IJzendoorn, Speranza, & Tambelli, 2000; Bakermans-Kranenburg & van IJzendoorn, 2009; Scharf, 2001). Because there were few adolescents in the insecure groups, we combined these groups into one insecure group and examined secure versus insecure AAI group differences. Adolescents' AAI group classifications were not linked to their Shipley vocabulary scores, $t(173) = 1.56$, $p = .12$.

ERCS and conflict behavior scores. We present descriptive statistics for participants' initial and follow-up ERCS scores in Table 2.

Test of Hypotheses Related to Reconstructive Memory for Adolescent–Parent Conflict

To test our hypotheses about attachment-related differences in reconstructive memory for adolescent–parent conflict, we conducted a series of general linear mixed model analyses using the Proc Mixed procedure in SAS/STAT® software (Version 9.1). In these analyses, the repeated factor was participants' T1 and T2 ERCS summary scores. The fixed factors were the time at which participants completed the ERCS (i.e., ERCS Time; two levels: T1 vs. T2), Reporting Family Member (four levels: adolescent reporting about teen–mother conflict, adolescent reporting about teen–father conflict, mother, and father), Adolescent Gender (two levels:

Table 2
Descriptive Statistics for Adolescents' and Parents' ERCS Scores

	Initial (T1)		Follow-up (T2)	
Variable	*M*	*SD*	*M*	*SD*
Adolescent ERCS scores for adolescent–mother conflict				
Positive interaction	3.36	0.94	3.32	0.80
Negative interaction	1.79	0.76	1.65	0.68
Hostile treatment	1.29	0.57	1.27	0.51
Adolescent ERCS scores for adolescent–father conflict				
Positive interaction	3.29	0.83	3.25	0.83
Negative interaction	1.78	0.68	1.60	0.67
Hostile treatment	1.27	0.41	1.26	0.44
Mother ERCS scores for adolescent–mother conflict				
Positive interaction	3.24	0.90	3.23	0.83
Negative interaction	1.67	0.73	1.67	0.70
Hostile treatment	1.41	0.73	1.38	0.63
Father ERCS scores for adolescent–father conflict				
Positive interaction	3.37	0.79	3.31	0.73
Negative interaction	1.65	0.66	1.65	0.59
Hostile treatment	1.39	0.66	1.41	0.61

Note. ERCS initial (T1) *n*s range from 180 to 189. ERCS follow-up (T2) *n*s range from 151 to 160. ERCS = Emotional Response to Conflict Scale.

boy vs. girl), and Adolescent Attachment (two levels: secure vs. insecure). We also added a "family" random effect in our analyses to account for any correlations among ERCS scores obtained from participants within the same family. We included all main effects and first-, second-, and third-order interactions in our analyses.

The results of interest in these mixed model analyses were whether the ERCS Time × Attachment Group interactions (and accompanying post hoc analyses examining T1/T2 differences) were significant for each reporting family member and whether different patterns of gender-related results emerged with regard to adolescent boys and/or girls. To obtain these particular results, we tested the simple interaction effects of ERCS Time × Adolescent Attachment Group within both the three-way ERCS Time × Adolescent Attachment Group × Reporting Family Member interactions and the four-way ERCS Time × Adolescent Gender × Adolescent Attachment Group × Reporting Family Member interactions. Then, when significant simple ERCS Time × Adolescent Attachment Group interaction effects emerged and subsequent post hoc tests provided evidence that change did indeed occur (i.e., significant attachment-related differences emerged between participants' T1 and T2 ERCS scores), we plotted participants' ERCS scores and conducted accompanying t tests to describe *how*, on average, family members' perceptions of conflict changed over a span of 6 weeks as a function of adolescent attachment and/or adolescent gender.

Moreover, for each of these examined significant interactions, we conducted further follow-up analyses to determine how participants' perceptions of the conflict converged or diverged from the behavioral ratings of the adolescent–parent conflict interactions over the 6-week period. These follow-up analyses mirrored the mixed model analyses described earlier, except that the repeated factor was two difference scores that represented the difference between participants' ERCS summary scores and our behavioral rating scores (one score for T1 perceptions and another for T2 perceptions). We calculated these difference scores by first standardizing participants' ERCS summary scores and behavioral hostility scores (behavioral conflict scores had already been standardized during the PCAs). We then subtracted each participant's behavior scores from his or her corresponding ERCS summary scores (i.e., ERCS positive interaction score − positive adolescent–parent dyadic conflict behavior scores; ERCS negative interaction score − positive adolescent–parent dyadic conflict behavior scores [reverse scored]; ERCS hostile treatment score − behavioral hostility score). Thus, for positive interaction, a positive (i.e., nonnegative) difference score demonstrated that the reporting family member perceived the conflict more favorably than did observers, and a negative (i.e., nonpositive) difference score demonstrated that the reporting family member perceived the conflict in a less favorable way than did observers. In contrast, for negative interaction and hostile treatment, a positive (i.e., nonnegative) difference score demonstrated that the reporting family member perceived the conflict less favorably than did observers, and a negative (i.e., nonpositive) difference score demonstrated that the reporting family member perceived the conflict in a more favorable way than did observers. Using these mixed model analyses and accompanying t tests, we were able to determine whether (a) difference scores were significantly different from zero (i.e., a significant difference score indicated that the reporting family members' perceptions and our behavioral ratings were significantly different), and (b) whether corresponding T1- and T2-related *difference scores* differed significantly (i.e., a significant difference in the difference scores indicated that that there was a significant change in the degree to which the T2 and T1 scores diverged from observer ratings provided at T1). In the next two sections, we present our results by hypothesis. In light of Roisman, Fraley, and Belsky's (2007) recent suggestions that individual differences in the AAI should also be examined continuously, we examined whether and how adolescents' reliable AAI coherence of mind scores were linked to adolescent and parent reconstructive memory for conflict. As expected, these results (which are available from M. J. Dykas) were substantively similar to the AAI group classification results reported in this article.

> *Hypothesis 1: Adolescent attachment security is linked to more positively and less negatively biased adolescent reconstructive memory for adolescent–parent conflict over time, whereas adolescent attachment insecurity is linked to less positively and more negatively biased memory over time.*

With respect to *adolescent–mother conflict*, a significant ERCS Time × Attachment simple interaction effect emerged for adolescents' ERCS negative interaction scores, $F(3, 701) = 5.23$, $p < .01$, but not ERCS positive interaction, $F(3, 699) = 1.90$, $p = .12$, or hostile treatment scores, $F(3, 767) = 1.61$, $p = .18$: Secure adolescents showed a significant decrease in their propensity to remember the conflict as negative, $M_{T1-T2} = 0.21$, $SE = 0.06$, $t_{T2-T1}(611) = 3.70$, $p < .001$, $g = .14$ (see Figure 1). For this t test and all other t tests reported in this section, the degrees of freedom were derived from the omnibus mixed model analysis. Moreover, the measure of effect size is reported as Hedge's g (see Cooper & Hedges, 1994). In contrast, insecure adolescents' memory for the negative features of the conflict remained stable during the 6-week period, $M_{T1-T2} = 0.03$, $SE = 0.08$, $t_{T2-T1}(598) = 0.37$, $p = .71$, $g = .02$; see also Figure 1. Inspection of the difference scores between secure adolescents' T1 and T2 ERCS negative interaction scores and their reverse-scored positive adolescent–mother dyadic conflict behavior scores indicated that secure adolescents' perceptions of the negative features of the conflict were initially significantly less negative than the observed behavioral ratings at T1, $M_{DiffT1} = -1.41$, $SE = 0.10$, $t(503) = -13.39$, $p < .001$, $g = .63$, and continued to diverge significantly and become even less negative than the behavioral ratings 6 weeks later at T2, $M_{DiffT2} = -1.73$, $SE = 0.11$, $t(499) = -15.31$, $p < .001$, $g = .70$; $M_{DiffT1-DffT2} = 0.32$, $SE = 0.08$, $t_{T1-T2}(572) = 3.74$, $p < .001$, $g = .17$. Insecure adolescents' perceptions of the negative features of the conflict were also initially significantly less negative than the behavioral ratings at T1, $M_{DiffT1} = -1.59$, $SE = 0.15$, $t(503) = -10.65$, $p < .001$, $g = .47$, and T2, $M_{DiffT2} = -1.63$, $SE = 0.16$, $t(476) = -10.38$, $p < .001$, $g = .47$, but did not diverge significantly from our behavioral ratings over the 6-week period, $M_{DiffT1-DffT2} = 0.04$, $SE = 0.12$, $t_{T1-T2}(562) = 0.36$, $p = .72$, $g = .01$.

With respect to *adolescent–father conflict*, a significant ERCS Time × Attachment simple interaction effect emerged for both adolescents' ERCS positive interaction scores, $F(3, 707) = 3.49$, $p < .05$, and ERCS negative interaction scores, $F(3, 709) = 4.91$, $p < .01$, but not for adolescents' ERCS hostile treatment scores, $F(3, 773) = 0.20$, $p = .90$. Regarding adolescents' positive interaction scores, secure adolescents' memory remained stable during the 6-week period, $M_{T1-T2} = 0.03$, $SE = 0.06$, $t_{T1-T2}(633) = 0.45$, $p = .65$, $g = .02$, whereas insecure adolescents showed a significant decrease in their propensity to remember the conflict as positive, $M_{T1-T2} = 0.21$, $SE = 0.09$, $t_{T1-T2}(620) = 2.46$, $p < .05$, $g = .09$ (see Figure 2). Inspection of the difference scores between secure adolescents' T1 and T2 ERCS positive interaction scores and their positive adolescent–father dyadic conflict behavior scores indicated that secure adolescents' perceptions of the positive features of the conflict did not differ from the observed behavioral ratings at T1, $M_{DiffT1} = -0.01$, $SE = 0.10$, $t(508) = -0.05$, $p = .96$, $g = .01$, or T2, $M_{DiffT2} = -0.05$, $SE = 0.11$, $t(502) = -0.44$ $p = .66$, $g = .02$, and did not diverge from the behavioral ratings over time, $M_{DiffT1-DffT2} = 0.04$, $SE = 0.08$, $t_{T1-T2}(617) = 0.58$, $p = .56$, $g = .02$. Insecure adolescents' perceptions of the positive features of the conflict also did not differ significantly from the observed behavioral ratings at T1, $M_{DiffT1} = 0.20$, $SE = 0.15$, $t(502) = 1.36$, $p = .17$, $g = .06$, or T2, $M_{DiffT2} = -0.06$, $SE = 0.16$, $t(472) = -0.39$, $p = .70$, $g = .02$, but there was a significant change in insecure adolescents' perceptions with

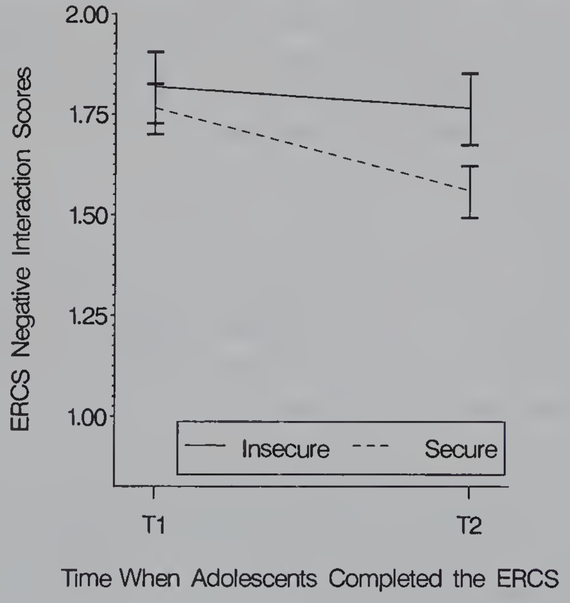


Figure 1. Adult Attachment Interview–related changes in adolescents' Emotional Response to Conflict Scale negative interaction scores for their conflict discussions with mother.

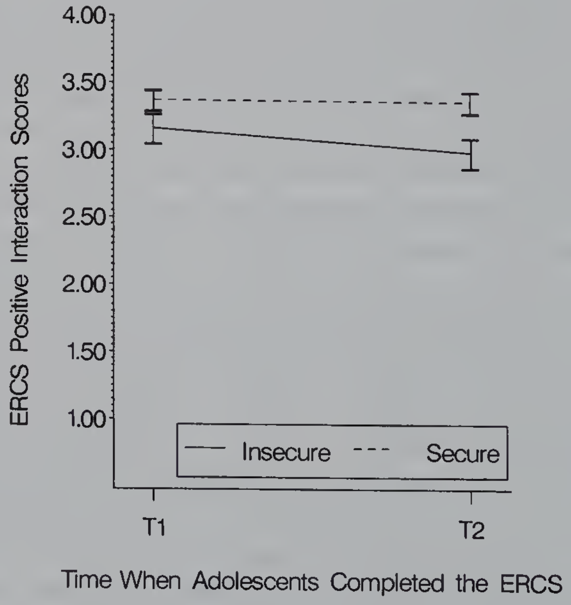


Figure 2. Adult Attachment Interview–related changes in adolescents' Emotional Response to Conflict Scale positive interaction scores for their conflict discussions with father.

respect to the behavioral ratings over time. More precisely, insecure adolescents' ratings shifted from being more positive to less positive over the 6-week period compared with the behavioral ratings, $M_{\text{DiffT1–DffT2}} = 0.26$, $SE = 0.10$, $t_{\text{T1–T2}}(607) = 2.55$, $p < .01$, $g = .11$.

Regarding adolescents' ERCS negative interaction scores for father, secure adolescents showed a significant decrease in their propensity to remember the conflict with their fathers as negative, $M_{\text{T1–T2}} = 0.17$, $SE = 0.06$, $t_{\text{T1–T2}}(617) = 2.96$, $p < .01$, $g = .11$, whereas insecure adolescents' memory for the negative features of the conflict remained stable during the 6-week period, $M_{\text{T1–T2}} = 0.15$, $SE = 0.08$, $t_{\text{T1–T2}}(601) = 1.92$, $p = .06$, $g = .08$ (see Figure 3). Inspection of the difference scores between secure adolescents' T1 and T2 ERCS negative interaction scores and their reverse-scored positive adolescent–father dyadic conflict observed behavior scores indicated that secure adolescents' perceptions of the negative features of the conflict were significantly less negative than the behavioral ratings at T1, $M_{\text{DiffT1}} = -1.53$, $SE = 0.11$, $t(515) = -14.23$, $p < .001$, $g = .61$, and later diverged significantly and became even less negative than the behavioral ratings 6 weeks later at T2, $M_{\text{DiffT2}} = -1.77$, $SE = 0.12$, $t(517) = -15.36$, $p < .001$, $g = .65$; $M_{\text{DiffT1–DffT2}} = 0.24$, $SE = 0.09$, $t_{\text{T1–T2}}(578) = 2.80$, $p < .01$, $g = .11$. Insecure adolescents' perceptions of the negative features of the conflict were also initially significantly less negative than the observed behavioral ratings at T1, $M_{\text{DiffT1}} = -1.73$, $SE = 0.15$, $t(510) = -11.46$, $p < .001$, $g = .51$, and T2, $M_{\text{DiffT2}} = -1.94$, $SE = 0.16$, $t(485) = -12.21$, $p < .001$, $g = .55$, but did not diverge significantly farther from the behavioral ratings over time, $M_{\text{DiffT1–DffT2}} = 0.21$, $SE = 0.12$, $t_{\text{T1–T2}}(565) = 1.80$, $p = .07$, $g = .07$.

> *Hypothesis 2: Adolescent attachment security is linked to more positively and less negatively biased parental reconstructive memory for adolescent–parent conflict over time, whereas adolescent attachment insecurity is linked to more negatively and less positively biased parental memory over time.*

With respect to *mothers*, a pair of contrasting gender-related ERCS Time × Attachment simple interaction effects emerged for mothers' ERCS positive interaction and hostile treatment scores, but not for mothers' ERCS negative interaction scores, $F(3, 700) = 2.08$, $p = .10$. With respect to mothers' ERCS positive interaction scores, a significant ERCS Time × Attachment simple interaction effect emerged for mothers of *boys* only, $F(3, 698) = 2.75$, $p < .05$: Whereas mothers of secure boys' memory for the positive features of the conflict remained stable during the 6-week period, $M_{\text{T1–T2}} = -0.05$, $SE = 0.10$, $t_{\text{T1–T2}}(627) = -0.49$, $p = .63$, $g = .02$, mothers of insecure boys showed a decrease in their propensity to remember the conflict as positive, $M_{\text{T1–T2}} = 0.37$, $SE = 0.13$, $t_{\text{T1–T2}}(620) = 2.82$, $p < .01$, $g = .11$ (see Figure 4). Inspection of the difference scores between mothers of secure boys' T1 and T2 ERCS positive interaction scores and their positive adolescent–mother dyadic conflict behavior scores indicated that secure boys' mothers' perceptions of

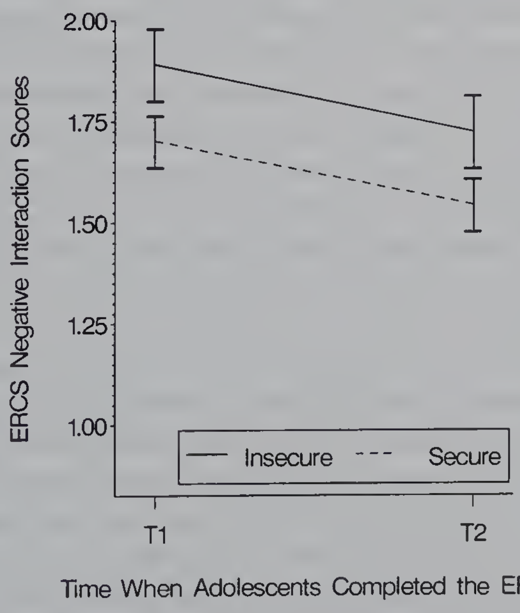


Figure 3. Adult Attachment Interview–related changes in adolescents' Emotional Response to Conflict Scale negative interaction scores for their conflict discussions with father.

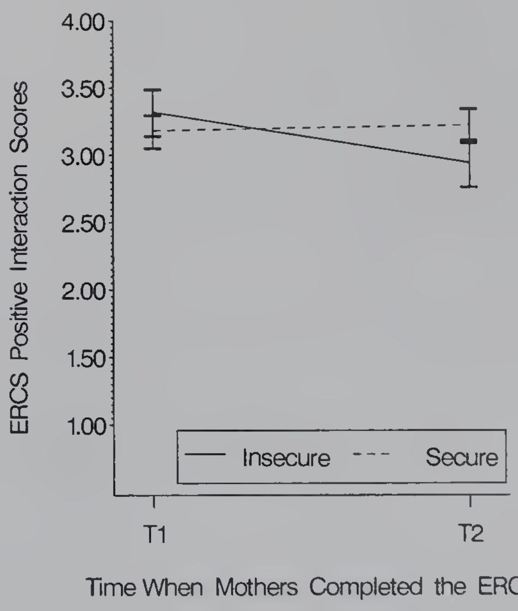


Figure 4. Adult Attachment Interview–related changes in mothers' Emotional Response to Conflict Scale positive interaction scores for their conflict discussions with sons.

the positive features of the conflict did not differ significantly from the behavioral ratings at T1, $M_{\text{DiffT1}} = -0.22$, $SE = 0.16$, $t(494) = -1.33$, $p = .18$, $g = .06$, or T2, $M_{\text{DiffT2}} = -0.15$, $SE = 0.17$, $t(479) = -0.86$, $p = .39$, $g = .04$, and did not change over time in reference to the behavioral ratings, $M_{\text{DiffT1–DffT2}} = -0.07$, $SE = 0.12$, $t_{\text{T1–T2}}(611) = -0.57$, $p = .57$, $g = .02$. Insecure boys' mothers' perceptions of the positive features of the conflict also did not differ significantly from observed behavioral ratings at T1, $M_{\text{DiffT1}} = 0.12$, $SE = 0.23$, $t(494) = 0.51$, $p = .61$, $g = .02$, or T2, $M_{\text{DiffT2}} = -0.33$, $SE = 0.24$, $t(464) = -1.35$, $p = .18$, $g = .06$, but the change in insecure boys' mothers perceptions of the conflict—from more to less positive—over the 6-week period was significant compared with the behavioral ratings, $M_{\text{DiffT1–DffT2}} = 0.44$, $SE = 0.16$, $t_{\text{T1–T2}}(606) = 2.79$, $p < .01$, $g = .11$.

In contrast, with respect to mothers' ERCS hostile treatment scores, a significant ERCS Time × Attachment simple interaction effect emerged for mothers of *girls* only, $F(3, 698) = 2.75$, $p < .05$: Mothers of secure girls showed a decrease in their propensity to remember being treated with hostility by their daughters during the conflict, $M_{\text{T1–T2}} = 0.15$, $SE = 0.06$, $t_{\text{T1–T2}}(645) = 2.36$, $p < .05$, $g = .10$, whereas mothers of insecure girls' memory for being treated with hostility during the conflict remained stable during the 6-week period, $M_{\text{T1–T2}} = -0.02$, $SE = 0.09$, $t_{\text{T1–T2}}(639) = -0.22$, $p = .83$, $g = .01$ (see Figure 5). Inspection of the difference scores between mothers of secure girls' T1 and T2 ERCS hostile treatment scores and their daughters' hostile behavioral scores indicated that although these mothers' perceptions of being treated with hostility did not differ from the observed behavioral ratings of daughter hostility at T1, $M_{\text{DiffT1}} = 0.11$, $SE = 0.12$, $t(767) = 0.87$, $p = .38$, $g = .03$, or T2, $M_{\text{DiffT2}} = -0.14$, $SE = 0.13$, $t(740) = -1.09$, $p = .27$, $g = .04$, the change from perceiving receiving more hostility to less hostility compared with our observations over time was significant, $M_{\text{DiffT1–DffT2}} = 0.25$, $SE = 0.11$, $t_{\text{T2–T1}}(623) = 2.32$, $p < .05$, $g = .09$. Insecure girls' mothers' perceptions of being treated with hostility by their daughters did not differ from the behavioral ratings at T1, $M_{\text{DiffT1}} = 0.30$, $SE = 0.18$, $t(781) = 1.67$, $p = .10$, $g = .06$, or T2, $M_{\text{DiffT2}} = 0.34$, $SE = 0.18$, $t(709) = 1.84$, $p = .07$, $g = .07$, and did not change over time in reference to the behavioral ratings, $M_{\text{DiffT1–DffT2}} = -0.04$, $SE = 0.15$, $t_{\text{T1–T2}}(619) = -0.27$, $p = .79$, $g = .01$.

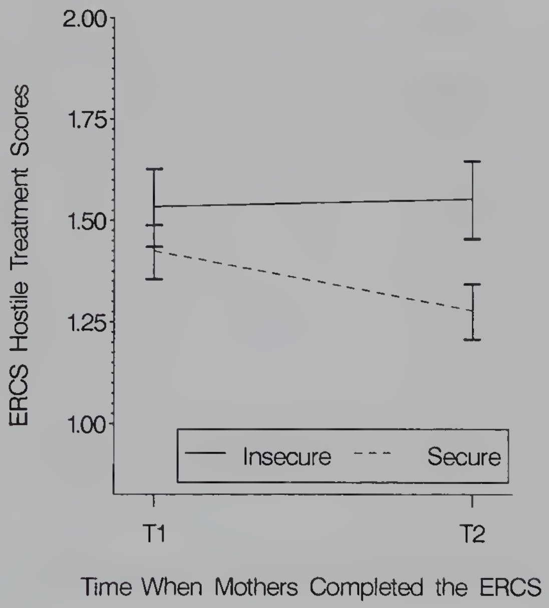


Figure 5. Adult Attachment Interview–related changes in mothers' Emotional Response to Conflict Scale hostile treatment scores for their conflict discussions with daughters.

For *fathers*, contrary to expectations, no significant ERCS Time × Attachment simple interaction effects or accompanying significant post hoc analyses examining T1–T2 differences emerged for adolescents' ERCS positive interaction, negative interaction, or hostile treatment scores (all post hoc *p*s ranged from .06 to .58).

Discussion

Given the importance of adolescent–parent conflict and evidence that memory for interpersonal experiences can lack accuracy over time (Baldwin, 1995; Fiske & Taylor, 1991), we examined attachment-related differences in the "reconstructions" of memory for adolescent–parent conflict and the ways in which these memory reconstructions were related to observer reports of the interactions. We found evidence that adolescents and their mothers, but not their fathers, reconstruct their perceptions of adolescent–parent conflict over time as a function of adolescent attachment, and we identified ways in which these reconstructions do and do not mesh with coder-rated observations of the conflict interaction.

First, findings suggest that when the memory for a specific adolescent–parent conflict degrades over a 6-week period, adolescents reconstruct their memories for the event as a function of their attachment organization. Thus, insecure adolescents, compared with secure adolescents, may have reported less favorable interactions over time because they were drawing on an insecure state of mind with respect to attachment to piece together the nature of these interactions. Overall, these attachment-related reconstructive memory biases fit well with and extend those that Feeney and Cassidy (2003) reported. However, unlike Feeney and Cassidy, we did not find links between our measure of attachment representations (the AAI) and every aspect of reconstructive memory for conflict with both mothers and fathers (i.e., positive interaction, negative interaction, and hostile treatment). Feeney and Cassidy, however, reported on links of adolescents' representations of a specific parent with memory about an interaction with that specific parent. We, on other hand, present findings indicating that a generalized, internal attachment organization (that is thought to be a consolidation of representations of mother, father, and other attachment figures; see Allen, 2008) relates to certain aspects of adolescents' memory for conflict with mother or father.

Moreover, in cases for which we found attachment-related memory biases, our analyses of difference scores between observer ratings and participant ratings of the interaction showed that secure adolescents had an initial bias to perceive their interactions with both mothers and fathers as *less* negative than observed scores, and this bias only grew stronger over time. These findings are particularly interesting because they suggest that secure adolescents are either not perceiving or not encoding the amount of negative content in their interactions with parents that observers perceive, and over time, this reporting bias is intensified because of their reliance on secure internal working models to aid memory recall. Insecure adolescents, on the other hand, showed a different pattern in their reports compared with observed scores of conflict. Insecure adolescents' initial ratings, like those of secure adolescents, were less negative than observer ratings, but there was no significant change over time in their ratings. Thus, although insecure adolescents perceived (or encoded) less negativity than did observers initially, they appeared to lack the protective memory bias that secure adolescents had that would allow them over time to recall the interaction as less negative. These findings illustrate the importance of considering comparisons between observers' assessments of behavior and self-reports over time to characterize the nature of attachment-based memory differences between secure and insecure adolescents.

It is interesting to note that although, as expected, relative to one another secure and insecure adolescents showed significantly distinct patterns of reconstructive memory, post hoc tests revealed findings that were not always expected. For instance, as expected, secure adolescents recalled their interactions with their mothers as less negative over time, yet contrary to expectation, insecure adolescents showed no significant change (rather than remembering their conflict interactions with their mothers as *more* negative over time). Research has shown that individuals tend to forget affect associated with unpleasant events more quickly than affect associated with pleasant events (Walker, Skowronski, & Thompson, 2003), and to minimize the affective experience of negative events (Taylor, 1991). The present results suggest that perhaps rather than having an overtly negative memory bias with respect to conflict interactions with mothers, insecure adolescents simply lack the normal protective memory process that secure adolescents use. (A similar lack of protective forgetting of negative affect emerged for insecure individuals in a study of counseling clients' memory for naturally occurring, in-session negative emotion: Specifically, clients high in attachment anxiety showed lower reductions in negative affect over the course of 1 week than did clients low in attachment anxiety; Woodhouse & Gelso, 2008.) In contrast to their interactions with their mothers, insecure adolescents *did* demonstrate a significant negative memory bias for interactions with their fathers, recalling the interaction as less positive than they had reported initially 6 weeks earlier. Secure adolescents, on the other hand, showed no significant change over time in their ratings of interaction with fathers. Moreover, the finding that secure versus insecure differences in adolescent reconstructive memory processes emerged in relation to negative emotion for mothers and positive emotion for fathers may indicate that reconstructive memory involves the salient emotional characteristics of these relationships (i.e., negativity is more frequent in adolescent–mother dyads, and positivity is more frequent in adolescent–father dyads; see Larson & Richards, 1994; Montemayor & Hanson, 1985; Smith & Forehand, 1986).

From an attachment perspective, attachment-related differences in reconstructive memory likely emerge because attachment organization serves as a structure through which new attachment-related information is processed rapidly and efficiently. These differences also likely stem from the proclivity of an individual's attachment organization to remain stable in the face of new information, especially information that would be inconsistent with previously obtained attachment-related knowledge (Bowlby, 1973). These notions are consistent with other theories in the social cognition literature, which state that internalized cognitive structures function to process social information in the most rapid and efficient ways possible, and in ways that are congruent with preexisting knowledge structures (Baldwin, 1992; Fiske & Taylor, 1991).

An additional important finding to emerge from this study was that adolescent AAI attachment security was linked not only to *adolescents'* reconstructive memory but also to *mothers'* reconstructive memory for conflict, such that compared with mothers of secure adolescents, mothers of insecure adolescents reconstructed aspects of the conflict interaction less favorably. Moreover, our follow-up analyses of how mothers of secure and insecure adolescents remembered their conflict interactions revealed expected patterns: Compared with mothers of secure boys, mothers of insecure boys typically remembered conflict interactions more negatively over time. Specifically, mothers of insecure boys initially rated the interaction no differently from the behavioral raters at T1 and T2; nevertheless, their memory for the positive aspects of the interaction became significantly less positive over time. Mothers of secure boys, on the other hand, likewise rated the positive aspects of the interaction no differently from the behavioral raters both at T1 and T2, yet did not show the significant decrease in their ratings of positivity that the mothers of insecure boys showed.

For mothers of girls, a somewhat different yet converging picture emerged. Mothers of secure girls reported less hostility over time as compared with mothers of insecure girls. Although neither T1 nor T2 ratings of daughters' hostility made by mothers of secure girls differed significantly from observer ratings of daughter hostility, there was a significant reduction in perceptions of hostility for mothers of secure girls. Mothers of insecure girls, on the other hand, likewise rated the hostile aspects of the interaction no differently from the behavioral raters both at T1 and T2, yet did not show the significant decrease in their ratings of hostility that the mothers of secure girls showed. Thus, mothers of insecure girls appeared to lack a protective ''forgetting'' of their daughters' hostility that mothers of secure girls seemed to demonstrate (although it is important to note that despite exhibiting an apparent protective ''forgetting,'' mothers of secure girls did not ultimately remember significantly less hostility than rated by observers).

These findings of maternal memory biases are noteworthy because they support attachment theorists' claims that parents' information-processing patterns are linked to their children's attachment security (see George & Solomon, 2008). It is striking that not only did we find compelling evidence with adolescents that is consistent with the idea that one's own state of mind with respect to attachment may guide one's memory (as predicted by attachment theory; Bowlby, 1969/1982; Main et al., 1985), we also found notable evidence of a link between one person's state of mind with respect to attachment and the social information processing of *another* person. If a mother has a memory bias that leads her to remember interactions less favorably over time, it is likely that she may feel differently about her adolescent child than she otherwise would. The adolescent's experience of having a mother who typically remembers interactions as more negative over time and feels more negatively about the adolescent may be a painful one. Moreover, such a negative maternal memory bias may lead the mother to feel, think, and act in ways that would interfere with the adolescent's tendency to view her as a secure base. Such a reduction in secure base experiences could contribute to the adolescent's overall insecure state of mind with respect to attachment. On the other hand, because the present study does not allow us to be certain about the direction of effects, it is important to consider that insecure adolescents might behave in ways that would make the mother less likely to remember positive aspects of the interaction. Regardless of the direction of the effect, it is notable that only mothers, and not fathers, showed a memory bias as a function of adolescent AAI. This finding is consistent with previous research suggesting that mothers may be more emotionally linked with their children than fathers (Larson & Richards, 1994; Montemayor, 1983; Youniss & Smollar, 1985). Further research will be necessary to better understand differences in the links between mother and father social information processing and adolescent attachment. Additionally, replication studies are needed to more fully understand the adolescent gender differences that emerged. For instance, mothers of secure adolescents reconstructed perceptions of girls' hostility as a function of girls' attachment but did not do the same with regard to boys' hostility. In most previous attachment research, gender has not moderated attachment-related findings (see Cassidy & Shaver, 2008, for reviews). It may be that hostility is more salient in mother–daughter relationships than in mother–son relationships, given empirical evidence that girls tend to engage in both greater conflict (as well as greater intimacy) with parents, particularly mothers, than do boys (Larson & Richards, 1994; Montemayor, 1983).

Although this study provides new insights into attachment-related memory processes, the reported findings should be considered within the context of the study's limitations. One limitation of the present study is that we were unable to examine links between parents' AAI classifications and reconstructive memory, an examination that might provide clues into possible pathways by which security of attachment can be transmitted within families across generations (see van IJzendoorn, 1995). It may be, for instance, that parental information processing mediates a (hypothesized) link between parent and adolescent attachment, such that once parent attachment is considered as a predictor of parental memory, the link between adolescent attachment and parental memory would be weakened (or no longer emerge). In other words, it may be that differences in mothers' attachment representations are driving differences in mothers' memory, mothers' behavior with the adolescent, and adolescent attachment.

Another limitation is that all families were maritally intact, limiting generalizability. Nevertheless, an important goal of the study was to examine memory biases for interactions occurring within relationships that adolescents have with both parents (see Collins & Russell, 1991, for a discussion of the importance of understanding the different contributions of mothers and fathers to development). Relatedly, our sample was nonclinical, relatively low risk, and relatively homogeneous with respect to socioeconomic status; memory processes may work differently in different kinds of samples. It is also important to note that our findings might have differed had we examined reconstructive memory over a different time span (e.g., over 6 months rather than 6 weeks); furthermore, we cannot rule out the possibility that the quality of adolescent–parent interactions after the laboratory visit contributed to how participants remembered the laboratory conflicts. An additional limitation is that we examined only secure versus insecure AAI group differences; future research (with samples containing larger numbers of insecure adolescents) is needed to reveal how adolescents with different AAI insecure classifications reconstruct memories for conflicts with parents. Finally, our results are correlational and do not allow us to determine causal relations. Although our interpretation of the data is consistent with attachment theory, the links that emerged in this investigation may have emerged for other reasons (e.g., insecurity may be associated with negative mood, and a wealth of data indicates that negative mood is linked to memory biases; see Haaga, Dyck, & Ernst, 1991).

In addition to addressing these limitations, future researchers could also examine reconstructive memory processes outside the context of adolescent–parent conflict (e.g., in contexts involving secure base support and safe haven provision), in other periods of development (e.g., childhood), and in relation to underlying psychophysiological processes (e.g., secure and insecure adolescents may employ different prefrontal emotion regulatory strategies when recalling past conflicts with parents). From a clinical perspective, these findings could have important implications for how adolescent–parent conflict is addressed during treatment. Instead of focusing solely on what occurs during an adolescent–parent conflict, clinicians might focus on how conflict is remembered over time as a function of adolescents' attachment security. By focusing on the attachment-related aftereffects of conflict, clinicians may be more resourceful in helping high-conflict adolescent–parent dyads overcome their difficulties.

References

Adams, R., & Laursen, B. (2007). The correlates of conflict: Disagreement is not necessarily detrimental. *Journal of Family Psychology*, *21*, 445–458.

Allen, J. P. (2008). The attachment system in adolescence. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 419–435). New York: Guilford.

Allen, J., Boykin-McElhaney, K., Kuperminc, G., & Jodl, K. (2004). Stability and change in attachment security across adolescence. *Child Development*, *75*, 1792–1805.

Allen, J. P., Moore, C., Kuperminc, G., & Bell, K. (1998). Attachment and adolescent psychosocial functioning. *Child Development*, *69*, 1406–1419.

Ammaniti, M., van IJzendoorn, M. H., Speranza, A. M., & Tambelli, R. (2000). Internal working models of attachment during late childhood and early adolescence: An exploration of stability and change. *Attachment and Human Development*, *2*, 328–346.

Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2009). The first 10,000 Adult Attachment Interviews: Distributions of adult attachment representations in clinical and non-clinical groups. *Attachment and Human Development*, *11*, 223–263.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*, 461–484.

Baldwin, M. W. (1995). Relational schemas and cognition in close relationships. *Journal of Social and Personal Relationships*, *12*, 547–552.

Belsky, J., Spritz, B., & Crnic, K. (1996). Infant attachment security and affective- cognitive information processing at age 3. *Psychological Science*, *7*, 111–114.

Benoit, D., Parker, K., & Zeanah, C. (1997). Mothers' representations of their infants assessed prenatally: Stability and association with infants' attachment classifications. *Journal of Child Psychology and Psychiatry*, *38*, 307–313.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books (Original work published 1969).

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation*. New York: Basic Books.

Bowlby, J. (1988). *A secure-base: Clinical applications of attachment theory*. London: Tavistock.

Bretherton, I., & Munholland, K. A. (1999). Internal working models in attachment relationships: A construct revisited. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 89–111). New York: Guilford.

Bretherton, I., & Munholland, K. A. (2008). Internal working models in attachment relationships: Elaborating a central construct in attachment theory. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 102–130). New York: Guilford.

Bugental, D. B., & Happaney, K. (2002). Parental attributions. In M. H. Bornstein (Ed), *Handbook of parenting: Vol. 3. Being and becoming a parent*. (2nd ed., pp. 509–535). Mahwah, NJ: Erlbaum.

Cassidy, J., (1998). *Emotional Response to Conflict Scale*. Unpublished instrument, University of Maryland, College Park.

Cassidy, J., & Shaver, P. R. (2008). *Handbook of attachment theory, research, and clinical applications*. New York: Guilford.

Collins, W. A., & Laursen, B. (1992). Conflict and relationships during adolescence. In C. U. Shantz & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 216–241). New York: Cambridge University Press.

Collins, W. A., & Russell, G. (1991). Mother–child and father–child relationships in middle childhood and adolescence: A developmental analysis. *Developmental Review*, *11*, 99–136.

Conger, R., & Ge, X. (1999). Conflict and cohesion in parent–adolescent relations: Changes in emotional expression from early to midadolescence. In M. J. Cox & J. Brooks-Gunn (Eds.), *Conflict and cohesion in families: Causes and consequences* (pp. 185–206). Mahwah, NJ: Erlbaum.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Ruseell Sage Foundation.

Dweck, C., & London, B. (2004, October). The role of mental representation in social development. *Merrill-Palmer Quarterly*, *50*, 428–444.

Dykas, M. J., & Cassidy, J. (2007). Attachment and the processing of social information in adolescence. In M. Scharf & O. Mayseless (Vol. Eds.) & R. Larson & L. A. Jensen (Series Eds.), *New directions for child and adolescent development. Attachment in adolescence: Reflections and new angles* (Vol. 117, pp. 41–56). San Francisco: Jossey-Bass.

Eisenberg, N., Hofer, C., Spinrad, T. L., Gershoff, E. T., Valiente, C., Zhou, Q., et al. (2008). Understanding mother-adolescent conflict discussions: Concurrent and across-time prediction from youths' dispositions and parenting. *Monographs of the Society for Research in Child Development*, *73*(2, Serial No. 290.

Feeney, B. C., & Cassidy, J. (2003). Reconstructive memory related to adolescent–parent conflict interactions: The influence of attachment-related representations on immediate perceptions and changes in perceptions over time. *Journal of Personality and Social Psychology*, *85*, 945–955.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Ge, X., Conger, R., & Elder, G. (1996). Coming of age too early: Pubertal influences on girls' vulnerability to psychological distress. *Child Development*, *67*, 3386–3400.

George, C., Kaplan, N., & Main, M. (1996). *Adult Attachment Interview protocol* (3rd ed.). Unpublished manuscript, University of California, Berkeley.

George, C., & Solomon, J. (2008). The caregiving system: A behavioral systems approach to parenting. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 833–856). New York: Guilford.

Grienenberger, J. F., Kelly, K., & Slade, A. (2005). Maternal reflective functioning, mother- infant affective communication, and infant attachment: Exploring the link between mental states and observed caregiving behavior in the intergenerational transmission of attachment. *Attachment and Human Development*, *7*, 299–311.

Grotevant, H. D. (1998). Adolescent development in family contexts. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 1097–1149). New York: Wiley.

Haaga, D., Dyck, M., & Ernst, D. (1991). Empirical status of cognitive theory of depression. *Psychological Bulletin*, *110*, 215–236.

Hesse, E.. (1999). The Adult Attachment Interview: Historical and current perspectives. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 395–433). New York: Guilford.

Hesse, E.. (2008). The Adult Attachment Interview: Protocol, method of analysis, and empirical studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research and clinical applications* (2nd ed., pp. 552–598). New York: Guilford.

Kirk, J., & Rattan, G. (1992). Critique of the Shipley Institute of Living Scale. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 9, pp. 490–495). Kansas City, MO: Test Corporation of America.

Kirsh, S. J., & Cassidy, J. (1997). Preschoolers' attention to and memory for attachment-relevant information. *Child Development*, *68*, 1143–1153.

Kobak, R. R., Cole, H. E., Ferenz-Gillies, R., Fleming, W. S., & Gamble, W. (1993). Attachment and emotion regulation during mother–teen problem-solving: A control theory analysis. *Child Development*, *64*, 231–245.

Larson, R., & Richards, M. H. (1994). *Divergent realities: The emotional lives of mothers, fathers, and adolescents*. New York: Basic Books.

Main, M., & Goldwyn, R. (1998). *Adult attachment rating and classification systems*. Unpublished manuscript, University of California, Berkeley.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. In I. Bretherton & E. Waters (Eds.), Growing points in attachment theory and research. *Monographs of the Society for Research in Child Development*, *50*(1–2, Serial No. 209), pp. 66–104.

Montemayor, R. (1983). Parents and adolescents in conflict: All families some of the time and some families most of the time. *Journal of Early Adolescence*, *3*, 83–103.

Montemayor, R., & Hanson, E. (1985). A naturalistic view of conflict between adolescents and their parents and siblings. *Journal of Early Adolescence*, *5*, 23–30.

Pasupathi, M. (2001). The social construction of the personal past and its implications for adult development. *Psychological Bulletin*, *127*, 651–672.

Roisman, G. I., Fraley, R. C., & Belsky, J. (2007). A taxometric study of the Adult Attachment Interview. *Developmental Psychology*, *43*, 675–686.

Russell, A., & Saebel, J. (1997). Mother–son, mother–daughter, father–son, and father–daughter: Are they distinct relationships? *Developmental Review*, *17*, 111–147.

Scharf, M. (2001). A "natural experiment" in childrearing ecologies and adolescents' attachment and separation representations. *Child Development*, *72*, 236–252.

Shipley, W. C. (1946). *Institute of Living Scale*. Los Angeles: Western Psychological Services.

Slade, A., Belsky, J., Aber, J. L., & Phelps, J. L. (1999). Mothers' representations of their relationships with their toddlers: Links to adult attachment and observed mothering. *Developmental Psychology*, *35*, 611–619.

Smetana, J. (1996). Adolescent–parent conflict: Implications for adaptive and maladaptive development. In D. Cicchetti & S. Toth (Eds.), *Adolescence: Opportunities and challenges* (pp. 1–46). Rochester, NY: University of Rochester Press.

Smetana, J., Campione-Barr, N., & Metzger, A. (2006). Adolescent development in interpersonal and societal contexts. *Annual Review of Psychology*, *57*, 255–284.

Smith, K. A., & Forehand, R. (1986). Parent–adolescent conflict: Comparison and prediction of the perceptions of mothers, fathers, and daughters. *Journal of Early Adolescence*, *6*, 353–367.

Strodtbeck, F. (1951). Husband–wife interaction over revealed differences. *American Sociology Review*, *16*, 463–473.

Taylor, S. E. (1991). Assymmetrical effects of positive and negative events: The mobilization–minimization hypothesis. *Psychological Bulletin*, *110*, 67–85.

Tversky, B., & Marsh, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, *40*, 1–38.

van Emmichoven, I. A. Z., van IJzendoorn, M. H., de Ruiter, C., & Brosschot, J. F. (2003). Selective processing of threatening information: Effects of attachment representation and anxiety disorder on attention and memory. *Development and Psychopathology*, *15*, 219–237.

van IJzendoorn, M. (1995). Adult attachment representations, parental responsiveness, and infant attachment: A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin*, *117*, 387–403.

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—And memory helps to keep it that way! *Review of General Psychology*, *7*, 203–210.

Woodhouse, S. S., & Gelso, C. J. (2008). Volunteer client adult attachment, memory for in-session emotion, and mood awareness: An affect regulation perspective. *Journal of Counseling Psychology*, *55*, 197–208.

Youniss, J., & Smollar, J. (1985). *Adolescent relations with mothers, fathers, and friends*. Chicago: University of Chicago Press.

Ziv, Y., Cassidy, J., & Ramos-Marcuse, F. (2002). *Adolescent–Parent Conflict Interaction coding system*. Unpublished manuscript, University of Maryland, College Park.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1460–1474

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood

Elizabeth Votruba-Drzal
University of Pittsburgh

Rebekah Levine Coley
Boston College

Carolina Maldonado-Carreño
Universidad de Los Andes

Christine P. Li-Grining
Loyola University of Chicago

P. Lindsay Chase-Lansdale
Northwestern University

Research examining the longer term influences of child care on children's development has expanded in recent years, but few studies have considered low-income children's experiences in community care arrangements. Using data from the Three-City Study ($N = 349$), the present investigation examines the influences of child care quality, extent and type on low-income children's development of behavior problems during middle childhood (7–11 years old). Higher levels of child care quality were linked to moderate reductions in externalizing behavior problems. High-quality child care was especially protective against the development of behavior problems for boys and African American children. Child care type and the extent of care that children experienced were generally unrelated to behavior problems in middle childhood.

Nonparental child care arrangements have become important contexts for early child development in the United States, with recent national estimates showing that approximately 12.2 million children under age 6 attend child care or preschool programs (Mulligan, Brimhall, & West, 2005). Economically disadvantaged children are slightly less likely than their advantaged counterparts to experience nonparental care during their preschool years, with 65% of children from poor families and 75% of children from nonpoor families regularly cared for in child care settings (Mulligan et al., 2005). Yet child care experiences may be particularly salient for low-income children's health and development. Faced with limited resources in their home environments, low-income children may be especially responsive either to the added supports of stimulating and responsive nonparental care or to the added risks of poor quality or extensive child care (Dearing, McCartney, & Taylor, 2009; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; Magnuson, Ruhm, & Waldfogel, 2007).

Beyond a few model early intervention programs and a handful of short-term longitudinal studies, our knowledge is limited concerning the implications of child care experiences for low-income children's later development. Whether child care experiences have long-term effects on low-income children's socioemotional functioning, such as their

---

We gratefully acknowledge the support of the following organizations: University of Pittsburgh's Central Research Development Fund's Small Grants Program. Government agencies: National Institute of Child Health and Human Development (R03 HD057294 "Child Care Resources in Low-Income Communities" and RO1 HD36093 "Welfare Reform and the Well-Being of Children"), Office of the Assistant Secretary of Planning and Evaluation, Administration on Developmental Disabilities, Administration for Children and Families, Social Security Administration, and National Institute of Mental Health. Foundations: The Boston Foundation, The Annie E. Casey Foundation, The Edna McConnell Clark Foundation, The Lloyd A. Fry Foundation, The Hogg Foundation for Mental Health, The Robert Wood Johnson Foundation, The Joyce Foundation, The Henry J. Kaiser Family Foundation, The W.K. Kellogg Foundation, The Kronkosky Charitable Foundation, The John D. and Catherine T. MacArthur Foundation, The Charles Stewart Mott Foundation, The David and Lucile Packard Foundation, The Searle Fund for Policy Research, and The Woods Fund of Chicago. A special thank you is also extended to the families who participated in the Three-City Study.

Correspondence concerning this article should be addressed to Elizabeth Votruba-Drzal, Department of Psychology, University of Pittsburgh, 210 South Bouquet Street, 4123 Sennott Square, Pittsburgh, PA 15260. Electronic mail may be sent to evotruba@pitt.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0011

skills in regulating their emotions and behaviors, getting along with peers, and refraining from inappropriate behaviors, is understudied. Behavioral and emotional skills are essential for early school success and also have long-term implications for children's mental health and behavioral functioning (Raver, 2002). As children transition into formal schooling and traverse middle childhood, lower problem behaviors and better behavioral skills are central indicators of developmental success with long-term repercussions for healthy functioning (Huston & Ripke, 2006). Indeed, problem behaviors and behavioral and emotional skills that are established in middle childhood exhibit significant continuity across adolescence and even into adulthood (Huston & Ripke, 2006).

Grounded in bioecological models of development (Bronfenbrenner & Morris, 1998; Sameroff, 1994), this study seeks to enhance our understanding of how experiences in early education and care settings shape the development of behavior problems among economically disadvantaged children into middle childhood. Bioecological theory describes human development as the result of proximal processes and reciprocal interactions between individuals and their environments. Driving future development are proximal processes that occur in microsystems, along with the characteristics of individuals that affect the contexts that they select into as well as the experiences that they elicit in these contexts. Within the child care microsystem, this model argues that caregiving experiences as well as children's individual characteristics together influence future developmental trajectories, with potentially divergent influences for children with different characteristics (Bronfenbrenner & Ceci, 1994; Bronfenbrenner & Morris, 1998).

Effects of Child Care Quality on Children's Behavior Problems and Socioemotional Functioning

Research to date has identified three aspects of child care experiences that are important potential influences on children's well-being: quality of care, extent of care, and type of care setting. High-quality child care is characterized by caregivers who are warm and responsive, who provide materials and experiences that stimulate learning and emotional support, and who afford consistency and structure. Experiences in high-quality care settings promote socioemotional development by helping children learn to regulate their emotions, behaviors, and attention; to get along with peers; and to comply with rules and requests. Indeed, research has proliferated documenting the associations between early education and care experiences and children's socioemotional development. In concurrent and short-term longitudinal studies with large, socioeconomically diverse samples, high-quality child care has been found to predict modestly lower behavior problems and negative affect as well as enhanced peer relations, sociability, and compliance (e.g., Love et al., 2003; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network [ECCRN], 1998, 2001). If children show improved regulation, affect, and peer skills in early childhood in relation to child care experiences, it is hypothesized that these skills will carry forward, promoting an easier transition into formal schooling and in turn leading to better social functioning later in childhood (National Research Council and Institute of Medicine, 2000). There is little existing evidence, however, that links between child care quality and early socioemotional functioning endure into middle childhood. Findings from the NICHD ECCRN (2005) and Belsky et al. (2007) suggest that child care quality from birth through preschool does not predict behavior problems or other measures of social functioning in middle childhood. Similarly, results from the Cost, Quality, and Outcomes Study found no links between quality of child care during preschool and socioemotional functioning in second grade, although significant links were identified between teacher–child closeness during early child care and lower behavior problems during middle childhood (Peisner-Feinberg et al., 2001).

There is a notable paucity of research assessing the effects of child care quality on economically disadvantaged long-term behavioral functioning. Although the studies discussed earlier include children from a range of socioeconomic backgrounds, most children were from middle-class families. In contrast to the experiences of more affluent children, economically disadvantaged families tend to face high levels of stress and have limited resources to invest in children, leading to less stimulating, responsive, and emotionally supportive home environments (Magnuson & Votruba-Drzal, 2009). Such economic disparities have implications for children's development of behavior problems (Duncan & Brooks-Gunn, 1997; Votruba-Drzal, 2006). High-quality nonparental care settings may protect economically disadvantaged children by providing them with opportunities and experiences that facilitate the mastery of central developmental challenges in early and middle childhood (Dearing et al., 2009).

Model early intervention programs, such as the Perry Preschool Project and the Abecedarian Program, show that high-quality early child care experiences, combined with comprehensive services for children and families, protect low-income children against significant behavioral problems later in life such as juvenile delinquency and incarceration (Campbell, Ramey, Pungello, Sparlin, & Miller-Johnson, 2002; Schweinhart et al., 2005). These interventions were administered by highly trained professionals and provided multifaceted services to low-income children and families. In contrast, recent research indicates that the quality of child care typically experienced by economically disadvantaged children is highly variable and, indeed, often rated as inadequate in meeting children's developmental needs (Fuller, Kagan, Loeb, & Chang, 2004; Li-Grining & Coley, 2006), questioning the generalizability of results from such model intervention programs.

Virtually no studies have considered whether the quality of child care available in low-income communities provides long-term protection against behavior problems for economically disadvantaged children. The study by Peisner-Feinberg et al. (2001) found the protective effects of teacher–child closeness on behavior problems to be sustained more strongly among children of mothers with low education, which was used as a proxy for low income. Assessing low-income children from the Three-City Study, Votruba-Drzal, Coley, and Chase-Lansdale (2004) found that the quality of child care experienced by preschool-age children (average age of 3 years) predicted improvements in socioemotional functioning over a 1½ year period, with higher quality care reducing internalizing and serious externalizing behavior problems and enhancing positive behaviors. Similar results were reported by Loeb, Fuller, Kagan, and Carrol (2004) using a different sample of low-income children. It is unknown whether such short-term protective effects of child care quality on low-income children's behavioral functioning extend to middle childhood.

Extent of Child Care Use and Children's Behavior Problems and Socioemotional Functioning

Along with child care quality, the extent of nonparental care that children are exposed to is a salient dimension of early experiences in child care. Increasing evidence suggests that extensive experience in nonparental care during the early years may be harmful for behavioral functioning. Early, extensive nonparental care may threaten behavioral development by undermining maternal sensitivity, as mothers have less time to spend with young children and become familiar with their signals (NICHD ECCRN, 1999), or by imposing high levels of stress on young children associated with extended peer exposure that taxes their nascent self-regulatory and social skills. In a series of studies using data from the NICHD Study of Early Child Care, the NICHD ECCRN (1998, 2003, 2005) and Belsky et al. (2007) found positive links between the quantity of care children experienced and externalizing behavior problems during early childhood, although these effects faded out by middle childhood. Negative links also emerged between the hours children were in care and their social competence; these results remained through the transition to middle childhood (age 8) but faded thereafter becoming nonsignificant by the end of middle childhood (age 12). It is important to note that in the NICHD study, child care characteristics were measured from age 3 months through 54 months. The extent of care that children experienced across early childhood was quite stable, such that children who experienced extensive care during infancy tended to do so consistently throughout early childhood (NICHD ECCRN, 2003). Thus, NICHD measures of the quantity of care tap into both early entry and extensive hours of child care, and research with this sample has not been able to determine whether negative effects are due to early entry, to extensive hours of care, or to some combination of these two factors.

It is unknown whether a detrimental effect of extensive nonparental care will emerge among low-income children. Recent work assessing extent of child care during preschool years (e.g., after infancy but prior to kindergarten) has found no negative effects for low-income children. A recent study using a nationally representative sample found no detrimental effects of high hours in center care for low-income children's behavioral functioning in kindergarten, although negative effects were apparent among middle- and high-income children (Loeb et al., 2007). Similarly, longitudinal studies with low-income samples suggest no effects of extent of care (Love et al., 2003; Votruba-Drzal et al., 2004). One recent study of Canadian children found that an earlier age of onset of nonparental care (prior to 9 months) was protective against the development of externalizing behavior problems for children of mothers with low education, although the extent of care per week was not a significant predictor of behavior problems (Côté et al., 2007).

Type of Care and Children's Behavior Problems and Socioemotional Functioning

A third important aspect of early care experiences is captured by the type of care children experience, that is, care provided in child care centers or prekindergarten programs versus day care homes or informal care arrangements with relatives or babysitters. Formal child care centers typically have more educated care providers and higher developmental quality. Home arrangements, on the other hand, on average have notably smaller group size and lower child–adult ratios, often include relatives of the child as the care provider, and provide more flexible and accessible care to families (Coley, Li-Grining, & Chase-Lansdale, 2006; Li-Grining & Coley, 2006). Although beneficial for academic development, recent evidence with large and nationally representative samples suggests that center-based care is linked to elevated levels of externalizing problems in early childhood and that these associations are evident through middle childhood (Belsky et al., 2007; Magnuson et al., 2007). Although some argue that center-based care is less developmentally appropriate for infants in particular in comparison with home-based arrangements (Loeb et al., 2007), negative effects of center care on behavioral functioning have emerged when considering center care use in the preschool years (e.g., Magnuson et al., 2007) or including both infancy and preschool (e.g., Belsky et al., 2007). Although there is little research specifying the mechanisms of these effects, center-based care arrangements, with larger group sizes, fewer adults, and frequent staff changes, may challenge young children's self-regulatory skills and provide greater opportunities for negative peer interactions (Dowsett, Huston, Imes, & Gennetian, 2008). Recent studies have further argued that these harmful effects of center care on children's behaviors are stronger among poor children in comparison with their more advantaged counterparts (Loeb et al., 2007; Magnuson et al., 2007). It is interesting to note, however, that when studies are able to separate effects of child care quality, extent, and type, no detrimental effects of center care have been found for low-income preschoolers' behavioral functioning (Loeb et al., 2004; Votruba-Drzal et al., 2004). It is unclear whether such effects might emerge later during middle childhood.

Does Child Care Quality Matter More for Certain Children?

Bioecological theories of development suggest that individual characteristics may interact with environmental experiences to influence future development (Bronfenbrenner & Morris, 1998; Sameroff, 1994). In the realm of child care, we hypothesize that the effects of early experiences in nonparental care settings may differ based on characteristics of the developing child, such as gender and race/ethnicity. A variety of factors may lead to differential effects of early child care experiences on behavior problems related to child gender. These factors include gender differences in the development of inhibitory control (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996) and in the nature of children's social interactions in same-sex peer groups (Fabes, Hanish, & Martin, 2003; Maccoby, 1998), as well as boys' overall greater vulnerability and reactivity to psychosocial stress (Zaslow & Hayes, 1986). If boys exhibit lower inhibitory control or greater peer conflict or show greater reactivity to stress, then the environmental challenges and opportunities imposed by nonmaternal care may have elevated effects on boys' psychosocial functioning.

A handful of recent studies have considered whether characteristics of nonpaternal care settings (e.g., quality, quantity, and type) are differentially related to children's behavioral functioning for girls versus boys. Most recent research has found no evidence of gender moderation in either the short or long term (Belsky et al., 2007; Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Peisner-Feinberg et al., 2001). On the other hand, recent research with low-income children from the Three-City Study found that high-quality child care appeared more protective for boys than girls when it came to the development of externalizing behavior problems, and low-quality care seemed to be especially risky for boys' development of serious internalizing behavior problems (Votruba-Drzal et al., 2004).

Another child characteristic that may moderate associations between early nonmaternal care experiences and the development of behavior problems in middle childhood is child race/ethnicity. Conceptual models highlight the importance of considering child care settings within the context of children's ethnic and cultural backgrounds, taking into account the ecological contexts of home and care settings (Johnson et al., 2003). A number of studies have reported race/ethnicity differences in the use of child care, with African American families more likely to access center-based care and Hispanics more likely to use informal kin care (Fuller, Holloway, & Liang, 1996; Radey & Brewster, 2007). Only recently has research begun to ask explicitly whether characteristics of child care settings (e.g., quality, type, and quantity) have differential effects on children's development across sociocultural contexts.

Two contrasting hypotheses have been proposed regarding child care effects across children from socioculturally diverse backgrounds. One argument is that high-quality child care can have an enhanced protective role for children experiencing risks from factors such as poverty, racial discrimination, immigration, or non-English-speaking status that might limit children's and families' access to economic and social resources. Similarly, poor quality child care may be particularly detrimental for such children. A contrasting argument suggests that child care settings may not meet the culturally specific needs of children from ethnic minority families (Burchinal & Cryer, 2003). Or, our current measures of child care quality may not be valid for ethnic minority children. These arguments would suggest that links between child care quality and child well-being would be attenuated among specific groups of ethnic minority children (Burchinal & Cryer, 2003).

Assessing two large, longitudinal studies of child care quality (the NICHD study and the Cost, Quality, and Child Outcomes Study), Burchinal and Cryer (2003) found support for neither argument. Rather, their results suggest that highly used measures of child care quality are equally valid and reliable across African American, Hispanic, and White samples, and furthermore that quality is similarly predictive of children's cognitive and socioemotional skills during the preschool years (see also Burchinal et al., 2000). It is important to note, however, that their samples had numerous limitations, particularly in regard to the percentage of Hispanic families, leading to questions concerning generalizability. In contrast, a recent study using data from the Early Childhood Longitudinal Study's Kindergarten Cohort (ECLS–K) found that center-based care was not related to heightened behavior problems among English-proficient Hispanic children but was a particularly strong predictor of behavior problems for African American children in kindergarten (Loeb et al., 2007). Furthermore, a third recent study found a particular aspect of child care quality, relationship-focused care, to be predictive of less adaptive functioning for Hispanic children but not related to functioning among African American children (Owen, Klausli, Mata-Otero, & Caughy, 2008). Together, these results suggest that the interactions between child care characteristics and children's ethnic backgrounds are complex and deserving of continued careful attention.

Research Aims

The current investigation aims to strengthen understanding of how low-income children's early education and care experiences shape their development into middle childhood. First, we will examine whether there are long-lasting associations between the developmental quality, extent, and type of child care settings low-income children experience and the development of their behavior problems into middle childhood. Second, consistent with the bioecological model of child development, we will explore whether associations between dimensions of child care settings and low-income children's development vary according to two key child characteristics—gender and race/ethnicity.

Method

Data for this article were drawn from the Three-City Study, a longitudinal, multimethod study of the well-being of low-income children and families in the wake of welfare reform. Two components of the Three-City Study were used: the main survey and the Embedded Developmental Study (EDS). The main survey was conducted with a household-based, stratified random sample of about 2,400 low-income children (aged 0–4 or 10–14 years) and their primary female caregivers in low-income neighborhoods in Boston, Chicago, and San Antonio. Three survey waves were conducted, in 1999 (90% screening rate and 83% interview completion rate), 2000–2001 (88% retention rate), and 2005 (80% retention rate from Wave 1). During each wave of the main survey, mothers participated in an in-home interview and children aged 2 and older completed individualized cognitive achievement assessments. The EDS portion of the Three-City Study provided a more intensive view into the lives of the preschool-aged children (2–4 years) from the main survey. The goal of the EDS was to capture rich detail about children's primary caregivers and early environments, assessed through additional mother interviews and videotaped assessments of mother–child interactions, biological father interviews, and observations of child care settings and interviews with child care providers. Children were eligible for the child care component of the EDS if they were in child care for 10 hr or more per week: 49% of 737 EDS children in Wave 1 and 40% of 670 children in Wave 2 were eligible. After obtaining permission from the mother and the child care provider, children were observed in their primary child care setting for at least 2 hr, and the child care providers were interviewed. The response rate for the child care component of the EDS was 70% in Wave 1 and 73% in Wave 2.

The current article is based on 349 families who participated in a child care observation at either Wave 1 or 2. Comparisons of the children in our sample to the broader sample of 2- to 4-year-olds in the Three-City Study revealed few significant differences, with the exception that the children in our sample were more likely to be African American and less likely to have parents who were employed and had slightly higher levels of education. Among the 349 children in the analytic sample, 5% were missing Wave 1 child behavior problems measures, 19% were missing a child care observation at one wave, 16% were missing data on at least one covariate, and 21% were missing Wave 3 child behavior problems measures. Our analysis of missing data suggested that the data were missing at random. Missing data were imputed using multiple imputation by chained equations implemented in Stata 10 (StataCorp, College Station, TX) to create 10 complete data sets (Royston, 2004, 2005), which were then analyzed using *mim* commands in Stata. Based on the relative efficiency calculation by Rubin (1987), 10 imputations were deemed sufficient for the level of missing data in our study.

Analytic Approach

A major challenge when considering child care's influence on children's development involves disentangling whether child care characteristics truly enhance development or whether it is simply the case that more advantaged parents have children who are more developmentally advanced and also choose higher quality child care. Certain characteristics of children and families, such as race/ethnicity, income, parental education and employment, social support, and children's age and gender, influence parents' decisions regarding child care (Fuller et al., 1996; Singer, Fuller, Keiley, & Wolf, 1998). Most studies have controlled for a variety of family differences that might influence both family decision making and children's development to isolate less biased estimates of child care's influence. Since it is impossible to measure all important family characteristics, an additional strategy is to limit the influence of unmeasured variables statistically.

In the current analyses, associations between child care characteristics and the development of behavior problems in middle childhood were modeled using a longitudinal lagged regression model that is based on an accumulation of inputs framework, which has been articulated most clearly in the work of NICHD ECCRN and Duncan (2003) and Blau (1999). This model suggests that child *i*'s development at time *t* is an additive function of all child care, maternal, child, and household inputs to a child's development prior to that point in time. So, for example, as shown in Equation 1, behavior problems during middle childhood (Wave 3) are expressed as a function of child care characteristics, including the type of care arrangement, number of hours spent in care, and the quality of the care that children have experienced across Waves 1 and 2 of the study.

$$\begin{aligned}\text{Child outcomes}_{3i} = B_0 &+ B_1\,\text{child outcomes}_{1i}\\ &+ B_2\,\text{child care}_{1-2i}\\ &+ B_3\,\text{maternal}_{1-2i}\\ &+ B_4\,\text{child}_{1-2i} + \varepsilon_i. \qquad (1)\end{aligned}$$

To reduce the threat of selection bias posed by measured characteristics of children and families, a series of child, maternal, and household factors aggregated over Waves 1 and 2 of the survey were included in the models as covariates. Covariates included child age, gender, and race/ethnicity as well as maternal education, employment, and marital status. Each of these factors has been associated with characteristics of early care experiences and behavior problems in prior research. Thus, the failure to include these variables in our analysis may result in omitted variable bias. A Time 1 measure of the same child outcome that was being modeled as the dependent variable at Time 3 was included as an additional covariate in the model to reduce omitted variable bias further. The regression coefficients are thus interpreted as the effects of each independent variable on changes in behavior problems over time (Kessler & Greenberg, 1981). Including the Time 1 child outcome as a covariate allows us to control for unmeasured, time-invariant differences in children that were present at the first interview (Cain, 1975; Chase-Lansdale et al., 2003). In this model, unmeasured, time-varying characteristics of children that may be related to child care selection at Time 1 or 2 and children's development over time may continue to bias estimates of the relation between child care characteristics and children's development.

In addition to considering main effects of child-care experiences, we explored whether or not the effects of child care type, extent and quality vary as a function of a child's gender and race. This was done by adding interactions between child care characteristics and both child gender and race/ethnicity to our model followed by a series of post hoc regression analyses to compare the nonreference race/ethnic groups. All analyses were weighted with probability weights, which strengthen our ability to make inferences to our population of interest, which includes all children living in low-income neighborhoods in Boston, Chicago, or San Antonio in households with incomes < 200% of the poverty line.

Measures

Child care characteristics The global developmental quality of each child's care arrangement was measured during child care observations at Waves 1 and 2 using widely used and well-validated instruments. Center-based care arrangements were rated using the Early Childhood Environment Rating Scale–Revised (ECERS–R; Harms, Clifford, & Cryer, 1998). The quality of day care homes and informal home care arrangements was measured using the Family Day Care Rating Scale (FDCRS; Harms & Clifford, 1989). Fifteen percent of the observations were independently double-coded, with average intraclass correlations (ICC) at the subscale level of .90 in Wave 1 and .89 in Wave 2 for the ECERS–R and .98 and .99 for the FDCRS. The Arnett Scale of Provider Sensitivity (Arnett, 1989) was used to measure the emotional and behavioral relationships between the care providers and children in both center- and home-based care arrangements. The Arnett Scale (α_{T1} = .94 and α_{T2} = .92) supplemented items related to the teacher–child relationships on the ECERS and FDCRS, which focus more on supervision and discipline. Composite scores had average ICCs of .81 and .82 at Waves 1 and 2. A global child care quality composite (α_{T1} = .94 and α_{T2} = .91) was created at each wave by combining subscales from the ECERS or FDCRS with the composite score from the Arnett (Growing Up in Poverty Project, 2000; Votruba-Drzal et al., 2004). The total quality scores from Waves 1 and 2 were then averaged to reflect the average level of child care quality that children experienced across the first two waves of the study.

The extent of child care children experienced was measured using maternal reports of the number of hours per week that children were in child care at Waves 1 and 2 of the survey. These numbers were then averaged to create a composite measure of the extent of care that children experienced across Waves 1 and 2 of the survey. Child care type was coded with dummy variables indicating whether children experienced consistent center care across both waves of the survey, center care at one wave only, or consistent home-based care across Waves 1 and 2 (omitted).

Behavior problems Behavior problems were measured using mothers' reports on the age-appropriate version of the Child Behavior Checklist (CBCL; Achenbach, 1991, 1992; Achenbach & Rescorla, 2001). Analyses focused on the internalizing and externalizing behavior problems subscales. The CBCL internalizing scale (αs = .83–.87) focuses on anxiety, depression, withdrawal, and somatic complaints, whereas the externalizing scale (α_{T3}s = .90–.91) includes items related to aggression and rule breaking. Standardized (*T* scores) subscale scores from Wave 3 were used as dependent variables, with scores from Wave 1 included as covariates to reduce bias related to unobserved heterogeneity. We also included an indicator for whether the 2- to 3-year-old version or the 4- to 18-year-old version of the CBCL was used in Wave 1, to control for differences in instrumentation.

Child characteristics. Child characteristics included as covariates included child age at Wave 3 represented in months, and gender. Child race was represented with a series of dummy variables indicating whether the child is of Hispanic (omitted), non-Hispanic African American, or White/other non-Hispanic origin. Child characteristics were obtained via mother report.

Maternal and household characteristics. Several maternal and household characteristics averaged across Waves 1 and 2 were also included as covariates. Covariates included maternal education (greater than a high school education vs. a high school degree or less), employment (employed at least 10 hr per week vs. not), and marital status (married vs. not). Characteristics of children and mothers are presented in Table 1 along with descriptive information about their child care arrangements. Children in the analytic sample averaged 3 years of age in the first wave (range = 2–5 years) and averaged 9 years of age in the third wave (range = 7–11). They were primarily African American (55%) and Hispanic (37%), and most lived with single mothers (84%) with relatively low education. Bivariate correlations of the child care and behavior problems variables are available from the first author by request.

Results

Child Care Characteristics and Low-Income Children's Development in Middle Childhood

Results from lagged ordinary least squares regressions on children's behavior problems in middle childhood (i.e., at Wave 3) are presented in

Table 1

Child Care and Demographic Characteristics of the Sample (N = 349)

	M	SD
Child care		
Quality	4.30	1.25
Type of care[a]		
Stable center care	24.50	43.02
Ever center care	28.30	45.08
Home care	47.10	49.92
Hours per week	30.40	14.60
Child behavior problems T scores		
Internalizing Wave 1	51.67	10.16
Internalizing Wave 3	52.03	10.44
Externalizing Wave 1	51.92	10.28
Externalizing Wave 3	53.96	10.29
Child characteristics		
Age	110.94	9.90
Boy[a]	52.43	49.94
Race[a]		
White/other	6.26	24.23
African American	57.55	49.43
Hispanic	36.17	48.05
Maternal characteristics		
Education[a]		
High school education or less	47.21	49.92
Greater than high school	52.78	49.92
Employed[a]	60.60	37.19
Married[a]	21.27	37.15
CBCL 2- to 3-year-old version[a]	73.98	43.87

Note. CBCL = Child Behavior Checklist.

[a] Proportions.

Table 2

Main Effects Models Examining the Influence of Child Care Characteristics on the Development of Behavior Problems in Middle Childhood

	Internalizing		Externalizing	
	Coefficient	SE	Coefficient	SE
Child care				
Quality	−1.29	0.79	−2.16*	0.89
Type of care				
Stable center care	−1.97	2.31	−0.51	2.92
Ever center care	−0.42	1.95	1.47	1.85
Hours per week	−0.03	0.06	−0.04	0.06
Child characteristics				
Age	0.18	0.12	0.17	0.11
Race				
White/other	2.25	3.96	5.01†	2.70
African American	−3.86†	1.95	1.71	2.01
Boy	1.13	1.70	−0.96	1.56
Maternal characteristics				
Education				
Greater than high school	1.73	1.76	−0.37	1.62
Employed	0.96	2.25	−2.25	2.36
Married	−0.48	2.43	−1.66	2.18
CBCL 2- to 3-year-old version	2.94	2.71	6.06*	2.69
Child outcomes Wave 1	0.36***	0.08	0.29***	0.08

Note. CBCL = Child Behavior Checklist.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{***}p < .001$.

Table 2. Child care quality was linked to significant reductions in externalizing behavior problems over time. The magnitude of this association was modest in size. A 1 SD increase in the child care quality composite was linked to .26 of a standard deviation decline in externalizing behavior problems. The association between child care quality and internalizing behavior problems in middle childhood was also negative and of a relatively similar size (.15 of a standard deviation) but failed to reach conventional levels of statistical significance ($p = .10$). Controlling for the quality of care, neither the type of child care that children experienced nor the extent of care were significantly related to the development of behavior problems in middle childhood.

The relatively high correlation between stable center care and average quality in our data ($r = .47$) is consistent with the literature on child care for low-income families but raises concerns when the measures are analyzed simultaneously in a regression model. Thus, we ran an additional set of model specifications to consider whether the quality findings hold when the measure of child care type is excluded from the models and to confirm that we do not detect a significant effect of center care when child care quality is dropped from the model. When the measure of child care type was dropped from the regression models presented in Table 2, the negative effects of child care quality on both internalizing ($B = -1.62$, $t = -2.31$, $p < .05$) and externalizing ($B = -2.15$, $t = -2.81$, $p < .01$) behavior problems strengthened slightly. When child care quality was excluded from the models presented in Table 2, associations between child care type and both internalizing and externalizing problems remained nonsignificant at conventional levels, although stable center care did predict declines in internalizing problems at a level bordering significance ($B = -4.07$, $t = -2.08$, $p = .05$). This set of specifications helps strengthen our certainty that child care quality is the predominate factor in our models protecting against the development of behavior problems during middle childhood.

Child Characteristics as Moderators

The second set of analyses considered whether associations between child care characteristics and children's development varied as a function of child gender and race/ethnicity. Before delving into the results of these analyses, it is important to consider whether there are significant differences related to child gender and race/ethnicity when it comes to our primary variables of interest. There were no significant differences in average levels of behavior problems by gender or race/ethnicity at Wave 1 or 3. Nor were there significant differences in average child care characteristics by child gender. Characteristics of child care experiences were similar by race/ethnicity as well, with the exception that the average level of child care quality of children falling into the White/other race/ethnic category was significantly higher than the average level of quality for both African American and Hispanic children.

Child care characteristics and gender. Regression results indicated that child gender moderated the link between child care quality and children's functioning (see Table 3). Specifically, high-quality child care appeared to be especially protective for boys' development of internalizing behavior problems. The results for externalizing behavior problems suggest a similar pattern, but the interaction term did not reach statistical significance. A standard deviation increase in the child care quality composite was associated with .37 more of a standard deviation decline in internalizing behavior problems for boys when compared to girls. There was also some evidence to suggest that child care type showed different patterned links with boys' and girls' behavior problems, with a tendency for center care (i.e., one wave of center care and two waves of consistent center care) to be more protective for girls than boys when it came to the development of internalizing problems, when comparing center care to home-based care. This same pattern emerged for externalizing behavior problems, but this interaction was not statistically significant. Finally, consistent with the results of the main effects models, no significant associations were observed between the extent of child care and the development of behavior problems for boys or girls.

Child care characteristics and race/ethnicity. The association between child care quality and behavior problems in middle childhood also varied as a function of race/ethnicity. The results presented in Table 3 show that child care quality seemed to be especially protective against the development of behavior problems among African American children. Greater child care quality was related to significantly greater reductions in externalizing behavior problems for African American children than for both Hispanic and White/other children. Although this interaction between child care quality and child race/ethnicity did not reach statistical significance for internalizing behavior problems, the same pattern of results emerged.

Table 3

Interaction Models Examining the Influence of Child Care Characteristics on the Development of Behavior Problems in Middle Childhood

	Internalizing		Externalizing	
	Coefficient	*SE*	Coefficient	*SE*
Main effects				
Child care				
Quality	1.89	1.30	0.86	1.71
Type of care				
Stable center care	−10.56*	4.58	−5.77	5.90
Ever center care	−3.81	3.35	−4.22	3.44
Hours per week	0.04	0.08	−0.02	0.08
Child characteristics				
Age	0.18	0.12	0.18	0.12
Race				
White/other	−2.47	16.71	−1.45	14.48
African American	7.55	8.92	13.59	8.75
Boy	11.72	8.92	4.15	7.45
Maternal characteristics				
Education				
Greater than high school	1.51	1.70	0.27	1.60
Employed	1.27	2.28	−2.14	2.25
Married	−0.65	2.40	−2.61	2.30
CBCL 2- to 3-year-old version	3.07	2.62	6.23*	2.60
Child outcomes Wave 1	0.39***	0.08	0.29***	0.08
Gender × Child Care interactions				
Boy × Quality	−3.05*	1.43	−1.99	1.42
Boy × Stable Center	7.85†	4.58	4.94	4.71
Boy × Ever Center	6.65†	3.87	4.82	3.55
Boy × Hours	−0.04	0.13	0.03	0.13
Race × Child Care interactions				
African American × Quality[a]	−2.35	1.55	−3.13*	1.51
African American × Stable Center[a]	6.28	4.78	3.64	4.95
African American × Ever Center[a]	0.64	3.92	4.48	3.83
African American × Hours[a]	−0.10	0.11	−0.02	0.13
White/Other × Quality[b]	0.21	3.42	1.77	2.61
White/Other × Stable Center[b]	8.29	10.76	−0.08	8.26
White/Other × Ever Center[b]	−3.68	8.94	2.09	6.92
White/Other × Hours[b]	0.04	0.14	−1.31	0.16

Note. CBCL = Child Behavior Checklist.
[a] African American children compared to Hispanic children.
[b] White/other children compared to Hispanic children.
[†] $p < .10$. *$p < .05$. ***$p < .001$.

Discussion

Utilizing multimethod data from a representative sample of young children in low-income neighborhoods, this study extends existing research by assessing longer term effects of child care characteristics on the development of behavior problems through middle childhood. In this analysis, we asked whether three aspects of child care experiences predicted behavior problems: the global developmental quality of care, the extent of care, and care type. It is important to note that in this study, we assessed children's experiences in child care during their preschool years; the data did not include information on nonparental care during infancy.

The Importance of Child Care Quality

The most consistent results from this analysis highlight the importance of the quality of care in the reduction of problem behaviors. Using well-validated, standardized observational measures of the global developmental quality of the care provided by child care settings and controlling for the type and extent of care, we found that higher quality care was protective against the development of behavior problems in middle childhood. Children who attended more responsive, stimulating, and well-structured care settings than their peers during their preschool years showed reductions in externalizing behavior problems by mid-elementary school. The pattern of results was similar for internalizing behavior problems, although insufficiently precise to reach statistical significance. This means, of course, that the reverse was true as well: Children attending lower quality child care showed more elevated behavior problems than their peers by mid-elementary school.

These results extend a very limited and contradictory base of research concerning whether child-care quality shows longer term links with socioemotional functioning into middle childhood. Peisner-Feinberg et al. (2001), for instance, found positive effects of close teacher–child relationships but not global developmental quality on second graders' social functioning, whereas other research has unearthed longer term benefits of high-quality care on cognitive skills but not socioemotional functioning into middle childhood. For example, the NICHD study found that high-quality care from birth through age 5 predicted higher cognitive scores in third and fifth grades (Belsky et al., 2007; NICHD Early Child Care Research Network, 2005), but found no effects on behavior problems or socioemotional functioning among their notably more advantaged sample.

Why might our results differ so substantially from those derived from more economically advantaged samples? A risk and resilience framework may be useful in reconciling these differences (Masten, Best, & Garmezy, 1990). Children in our sample, primarily from ethnic minority, single-parent families with extremely limited economic resources and residing in high-poverty urban neighborhoods, likely faced an accumulation of risks across multiple contexts. Although most children have the resources to cope with one risk without serious developmental consequences, the cumulative risk perspective suggests that the accumulation of risk across settings leaves children vulnerable to maladaptive psychosocial functioning (Friedman & Chase-Lansdale, 2002). Faced with contextual risks from multiple sources, high-quality, responsive, and stimulating child care settings may serve as a protective factor, or source of resilience, for young children to avoid problematic behavior and to develop nascent emotional and behavioral skills in self-regulation, peer interactions, and conflict management during their early school years. Similarly, low-quality child care may exacerbate the existing risks children experience. Indeed, it is interesting to note that the current results show stronger and more consistent effects of child care quality on behavior problems than found by Votruba-Drzal et al. (2004) looking at short-term effects predicting children's well-being during early childhood.

A second set of core findings from this study indicated that particular subgroups of children were driving our findings related to child care quality. Based upon bioecological theory that suggests that the effects of contextual experiences may differ in response to characteristics of the individual, we hypothesized that child gender and race/ethnicity may moderate links between child care and children's development of behavior problems. Our results supported this hypothesis. Regarding gender, our findings suggest that higher quality early education and care experiences were particularly important for boys. Child care quality more strongly predicted a decline in boys' than girls' internalizing behavior problems, and a similar pattern of results was evident for externalizing problems although it was not statistically significant. The behavior problems of boys may be more responsive to the quality of their early nonmaternal care arrangements due to their greater challenges with self-regulation, wherein boys tend to have lower levels of inhibitory control and physiological regulation than girls (Dettling, Parker, Lane, Sebanc, & Gunnar, 2000; Kochanska et al., 1996). Moreover, during preschool, boys' peer interactions appear to incorporate more conflict and roughness than girls' play, with the latter involving more verbal interactions and cooperation (Fabes et al., 2003; Maccoby, 1998). Thus, as noted by Votruba-Drzal et al. (2004), more responsive and well-structured child care settings may provide an important support to boys in helping to regulate their behaviors and emotions and supervise their play in a productive manner.

In addition to this gender moderation, results also found that race/ethnicity was important. More specifically, child care quality was particularly protective against the development of behavior problems for African American children when compared to Hispanic and White/other children. When they were cared for in high-quality settings, African American children experienced greater declines in externalizing behavior problems than both Hispanic and White/other children. The same pattern of results emerged for internalizing problems, although the interactions were not statistically significant. These results may be related to the greater acceptance and wider use of formal child care and maternal employment among African American children, resulting in a better match between families' goals and expectations and child care settings. As noted by Johnson, Jaeger, Randolph, Cauce, Ward, and NICHD ECCRN (2003), in understanding ethnic minority families' experiences in child care, we must attend to their broader cultural and ecological contexts. Based upon theories of person-environment fit and models of competency in children of color (Garcia Coll et al., 1996; Johnson et al., 2003), this conceptualization argues that children's development must be understood in context, with minority children's contexts affected by their social and cultural location. Hispanic families may have fewer social norms concerning maternal employment and nonmaternal care provision; they also may experience greater disconnects between the language environment of child care settings and home settings (Magnuson & Waldfogel, 2005). Indeed, other research has found that Hispanic parents, particularly in immigrant families, are less likely to access a broad range of social services in addition to formal child care and these restrictions are linked with poorer outcomes for children (Kalil & Chen, in press).

If the context of the care setting provides a stronger fit with the family preferences and cultural norms of African American children, whereas language or cultural disconnects between home and care settings inhibit person–environment fit for Hispanic children, we would expect that supportive characteristics of care settings such as high-quality, structured, responsive care giving would have a stronger beneficial effect on African American children's development. Supporting the tenets of bioecological theory, this proposition argues that environmental experiences (e.g., child care) are both affected by and interact with individual characteristics of the child and family. Indeed, other research with the Three-City Study child care sample found that African American mothers reported greater satisfaction with their child's care setting than did Hispanic and White mothers, and also reported (nonsignificantly) higher levels of communication and comfort with the care provider (Coley et al., 2006). Future research should seek to explore these interactions more closely, for example assessing whether a match between the language or ethnicity of the care provider and the child improves the quality of the child's care experiences and enhances the beneficial effects of high-quality care or the detrimental effects of early entry or long hours in nonparental care settings on children's development.

It is interesting to note that the stronger beneficial effects of child care quality for African American than for Hispanic children found in the current research contrast with a lack of ethnic moderation found in larger scale studies assessed by Burchinal and Cryer (2003). As the authors remind us, those larger scale studies included extremely small samples of Latino families and excluded families who were not fluent in English. They also had small groups of low-income families. Although our sample is smaller, it is representative of low-income families in our three cities, and includes both English- and Spanish-speaking Hispanic children, with approximately 20% of children residing in immigrant families. It is possible that this more representative sample of low-income, ethnic minority children was better poised to unearth important ethnic differences in the effects of child care on children, over a longer time period. Given the paucity of research in this arena and the continuing notable expansion of ethnic diversity in the United States, continued assessment of ethnic and cultural variation in child-care effects on child well-being is clearly warranted.

Limited Results for Child Care Type and Extent

In addition to the quality of child care, this study also assessed the effect of child care type and the extent of use. As mentioned previously, studies using data from the NICHD Study of Early Child Care and Youth Development (SECCYD) and the ECLS–K have uncovered harmful effects of extensive or early-initiated child care on behavior problems, some of which endure in middle childhood (Belsky et al., 2007; Loeb et al., 2007). Similarly, these studies have found that the use of center-based care is linked with heightened behavior problems continuing into elementary school, albeit also with enhanced cognitive skills (Belsky et al., 2007; Loeb et al., 2007; Magnuson et al., 2007; NICHD ECCRN & Duncan, 2003). These patterns were not replicated in this study. Categorizing children's preschool experiences over two waves as being housed consistently in home-care settings, consistently in formal center care, or inconsistently in center care, and measuring the hours per week children experienced child care, we found that neither care type nor extent were linked to behavior problems in middle childhood. One exception to this pattern was a trend finding that suggested that center care in comparison with home-based care may be beneficial for girls but not boys when it comes to the development of internalizing problems.

In comparing these results to other longitudinal studies, it is important to note that we did not assess hours of care or center care starting in infancy but focused only on the preschool years. Extensive and large-group care settings may be more taxing for infants and toddlers than preschool-age children. Moreover, returning to the risk and resilience perspective noted earlier, it is possible that for children in high-poverty neighborhoods and families, the challenges of center care or extensive child care may not pose a measurable burden in isolation. Facing enhanced rates of residential moves and relationship transitions in comparison with advantaged families, poor children may react in a less discernable fashion to the extent and type of child care, with process variables like quality holding greater importance. Indeed, a previous shorter term study with the Three-City Study sample found that extent of care was linked with behavior problems only in interaction with the quality of care: More extensive experience in high-quality care predicted decreased behavior problems, whereas extensive experience in low-quality care predicted increases in behavior problems (Votruba-Drzal et al., 2004).

Conclusions

This study provides an important addition to existing research on child care arrangements and children's developmental trajectories. Overall, our results highlight the importance of the developmental quality of care experiences for disadvantaged young children, arguing that high-quality child care experiences can have a sustained influence on children's behavioral functioning, through their transition to formal schooling and into middle childhood. Early child care quality is important for understanding low-income children's ability to function effectively in school, get along with their peers, and control their impulses. Future research should seek to assess a broader array of behavioral and emotional functioning measures, for example, assessing prosocial behaviors and social skills.

This study has several strengths. One is the sampling of the Three-City Study, which provides an in-depth look at a representative sample of young children and their families living in economic disadvantage, at both the family and community levels, in three cities. As such, this sample, whose experiences of urban poverty place children at notable risk for maladaptive social functioning, represents a population of significant concern to policy makers and practitioners. Increasing understanding of what types of supportive services can enhance the well-being of such children is of paramount importance for practitioners and policy makers. Moreover, our study focuses on children and families' experiences in community care, which are the care settings that disadvantaged parents have been able to access in their communities. Although model early intervention studies have provided centrally important information on how to productively intervene and provide bundles of high-quality services to disadvantaged families, such studies cannot tell us whether their findings can be generalized to the types of services to which disadvantaged families actually have access.

It is also important to point out limitations of the current study. As in all nonexperimental work, we cannot draw causal conclusions from the results and remain cognizant that the links detected here, between child care environments and children's developmental trajectories, could be due to selection or other unmeasured variables. The inclusion of child and family characteristics as well as earlier child functioning lessens the likelihood of such alternative explanations. Still, we were not able to control for characteristics of important contexts, such as schools and homes, in the years between preschool and middle childhood. If these were correlated with child care characteristics, this would threaten the validity of our results. The analytic sample was also relatively small and representative of a particular population of disadvantaged urban families. Although attention was paid to carefully measuring characteristics of both child care centers and more informal home arrangements, this study was not able to compare children's experiences in maternal care to those of their peers in child care settings. Continued careful analysis of early child care environments and children's development using diverse samples and multiple methods that include measures of both behavior problems and social competence will help the field to triangulate evidence of how child care can support low-income children's healthy development. Further studies can also more carefully assess individual children's experiences in child care settings (rather than a more global classroom and group-level assessments obtained through measures such as the ECERS, FDCRS, and Arnett). Such research might shed light on the differential patterns related to child gender and ethnicity unearthed here.

Finally, studies have raised questions about the validity of the CBCL for high-poverty or ethnic minority children and the results of these studies are mixed. Some have validated the CBCL with low-income, African American, and Latino families (Gross et al., 2006; Sivan, Ridge, Gross, Richardson, & Cowell, 2008), whereas others have uncovered problems with using CBCL norms for these populations (Raadal, Milgrom, Cauce, & Mancl, 1994; Sandberg, Meyer-Bahlburg, & Yager, 1991). Thus, it is important for our findings to be replicated using alternative measures of behavior problems.

In sum, this article is part of a new generation of child care research that is taking a more nuanced approached to multiple dimensions of children's experiences in child care. With a sizable sample, valid, reliable, and longitudinal measures and rigorous statistical methodology, we have increasing confidence that child care quality is an important factor in the developmental trajectories of young children in poverty. The results of the present study add to a growing body of empirical evidence suggesting the need for policy and programmatic efforts to increase low-income families' access to high-quality child care.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington: University of Vermont Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*, 541–552.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., et al. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701.

Blau, D. (1999). The effect of child care characteristics on child development. *Journal of Human Resources*, *34*, 786–822.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*, 568–586.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993–1028). New York: Wiley.

Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly*, *18*, 401–426.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science*, *4*, 149–165.

Cain, G. G. (1975). Regression and selection models to improve nonexperimental comparisons. In C. A. Bernett & A. A. Lumsdiane (Eds.), *Evaluation and experiment* (pp. 297–317). New York: Academic Press.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparlin, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, *6*, 42–57.

Chase-Lansdale, P. L., Moffitt, R., Lohman, B. J., Cherlin, A., Coley, R. L., Pittman, L. D., et al. (2003). Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents. *Science*, *299*, 1548–1552.

Coley, R. L., Li-Grining, C., & Chase-Lansdale, P. L. (2006). Low-income families' child care experiences: Meeting the needs of children and families. In N. Cabrera, R. Hutchins, & E. Peters (Eds.), *From welfare to child care: What happens to children when mothers exchange welfare for work* (pp. 149–170). Mahwah, NJ: Erlbaum.

Côté, S. M., Boivin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., et al. (2007). The role of maternal education and non-maternal care services in the prevention of children's physical aggression. *Archives of General Psychiatry*, *64*, 1305–1312.

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does higher quality early child care promote low-income children's math and reading achievement in middle childhood? *Child Development*, *80*, 1329–1349.

Dettling, A. C., Parker, S. W., Lane, S. K., Sebanc, A. M., & Gunnar, M. R. (2000). Quality of care and temperament determine whether cortisol levels rise over the day for children in full-day child care. *Psychoneuroendocrinology*, *24*, 819–836.

Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly*, *23*, 69–93.

Duncan, G. J., & Brooks-Gunn, J. (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.

Fabes, R. A., Hanish, L. D., & Martin, C. L. (2003). Children at play: The role of peer in understanding the effects of child care. *Child Development*, *74*, 1039–1043.

Friedman, R., & Chase-Lansdale, P. (2002). Chronic adversities. In M. Rutter & E. Taylor (Eds.), *Child and adolescent psychiatry* (4th ed., pp. 261–276). Oxford, UK: Blackwell.

Fuller, B., Holloway, S., & Liang, X. (1996). Family selection of child-care centers: The influence of household support, ethnicity, and parental practices. *Child Development*, *67*, 3320–3337.

Fuller, B., Kagan, S. L., Loeb, S., & Chang, Y. (2004). Child care quality: Centers and home settings that serve poor families. *Early Childhood Research Quarterly*, *19*, 505–527.

Garcia Coll, C., Lamberty, G., Jenkins, R., McAdoo, H. P., Crnic, K., Wasik, B. H., et al. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development*, *67*, 1891–1914.

Gross, D., Young, M., Fogg, L., Ridge, A., Cowell, J., Richardson, R., et al. (2006). The equivalence of the Child Behavior Checklist/1½–5 across parent race/ethnicity, income level, and language. *Psychological Assessment*, *18*, 313–323.

Growing Up in Poverty Project. (2000). *Remember the child: Mothers balance work and child care under welfare reform*. Berkeley, CA: Graduate School of Education–PACE.

Harms, T., & Clifford, R. M. (1989). *The Family Day Care Rating Scale*. New York: Teachers College Press.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised*. New York: Teachers College Press.

Huston, A. C., & Ripke, M. N. (Eds.). (2006). *Developmental contexts in middle childhood: Bridges to adolescence and adulthood*. New York: Cambridge University Press.

Johnson, D. J., Jaeger, E., Randolph, S. M., Cauce, A. M., Ward, J., & NICHD ECCRN. (2003). Studying the effects of early child care experiences on the development of children of color in the United States: Toward a more inclusive research agenda. *Child Development*, *74*, 1227–1244.

Kalil, A., & Chen, J. (2008). Family citizenship status and food insecurity among low-income children of immigrants. *New Directions in Child and Adolescent Development.*, *121*, 43–62.

Kessler, R., & Greenberg, E. F. (1981). *Linear panel analysis: Models of quantitative change*. New York: Academic Press.

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*, 490–507.

Li-Grining, C. P., & Coley, R. L. (2006). Child care experiences in low-income communities: Developmental quality and maternal views. *Early Childhood Research Quarterly*, *21*, 125–141.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66.

Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *75*, 47–65.

Love, J. M., Harrison, L., Sagi-Schwartz, A., van IJzendoorn, M. S., Ross, C., Ungerer, J. A., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, *74*, 1021–1033.

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51.

Magnuson, K. A., & Votruba-Drzal, E. (2009). Enduring influences of childhood poverty. In S. Danziger & M. Cancian (Eds.), *Changing poverty* (pp. 153–179). New York: Russell Sage Foundation.

Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children*, *15*, 169–196.

Masten, A., Best, K., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology*, *2*, 425–444.

Mulligan, G. M., Brimhall, D., & West, J. (2005). *Childcare and early education arrangements of infants, toddlers, and preschoolers: 2001 (NCES 2006-039)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

National Research Council and Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, *69*, 1145–1170.

NICHD Early Child Care Research Network. (1999). Child care and mother–child interaction in the first 3 years of life. *Developmental Psychology*, *35*, 1399–1413.

NICHD Early Child Care Research Network. (2001). Child care and children's peer interaction at 24 and 36 months: The NICHD study of early child care. *Child Development*, 72, 1478–1500.

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, *74*, 976–1005.

NICHD Early Child Care Research Network. (2005). Early child care and children's development in the primary grades: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *43*, 537–570.

NICHD Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*, 144–1475.

Owen, M. T., Klausli, J. F., Mata-Otero, A., & Caughy, M. O. (2008). Relationship-focused child care practices: Quality of care and child outcomes for children in poverty. *Early Education and Development*, *19*, 302–329.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., Kagan, S., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72, 1534–1553.

Raadal, M., Milgrom, P., Cauce, A. M., & Mancl, L. (1994). Behavior problems in 5- to 11-year-old children from low-income families. *Journal of the American Academy of Child & Adolescent Psychiatry*, *33*, 1017–1025.

Radey, M., & Brewster, K. L. (2007). The influence of race/ethnicity on disadvantaged mothers' child care arrangements. *Early Childhood Research Quarterly*, *22*, 379–393.

Raver, C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report of the Society for Research in Child Development*, *16*, 1–20.

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, *4*, 227–241.

Royston, P. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal*, *5*, 527–536.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Sameroff, A. (1994). Developmental systems and family functioning. In R. D. Parke & S. G. Kellam (Eds.), *Exploring family relationships with other social contexts* (pp. 199–214). Hillsdale, NJ: Erlbaum.

Sandberg, D. E., Meyer-Bahlburg, H. F. L., & Yager, T. J. (1991). The Child Behavior Checklist nonclinical standardization samples: Should they be utilized as norms? *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 124–134.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40. (Monographs of the High/Scope Educational Research Foundation, 14)*. Ypsilanti, MI: High/Scope Press.

Singer, J. D., Fuller, B., Keiley, M. K., & Wolf, A. (1998). Early child-care selection: Variation by geographic location, maternal characteristics, and family structure. *Developmental Psychology*, *34*, 1129–1144.

Sivan, A., Ridge, A., Gross, D., Richardson, R., & Cowell, J. (2008). Analysis of two measures of child behavior problems by African American, Latino, and non-Latino Caucasian parents of young children: A focus group study. *Journal of Pediatric Nursing*, *23*, 20–27.

Votruba-Drzal, E. (2006). Economic disparities in middle childhood development: Does income matter? *Developmental Psychology*, *42*, 1154–1167.

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, *75*, 296–312.

Zaslow, M. J., & Hayes, C. D. (1986). Sex differences in children's responses to psychosocial stress: Toward a cross-context analysis. In M. Lamb & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol. 4, pp. 285–337). Hillsdale, NJ: Erlbaum.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1475–1489

Children's Understanding of Ordinary and Extraordinary Minds

Jonathan D. Lane, Henry M. Wellman, and E. Margaret Evans
University of Michigan

How and when do children develop an understanding of extraordinary mental capacities? The current study tested 56 preschoolers on false-belief and knowledge-ignorance tasks about the mental states of contrasting agents—some agents were ordinary humans, some had exceptional perceptual capacities, and others possessed extraordinary mental capacities. Results indicated that, in contrast to younger and older peers, children within a specific age range reliably attributed fallible, human-like capacities to ordinary humans and to several special agents (including God) for both tasks. These data lend critical support to an anthropomorphism hypothesis—which holds that children's understanding of extraordinary minds is derived from their everyday intuitive psychology—and reconcile disparities between the findings of other studies on children's understanding of extraordinary minds.

Children and adults worldwide come to understand persons as intentional agents who act in accordance with their perceptions, desires, and beliefs (Flavell & Miller, 1998; Wellman, 1990). Most people also come to entertain ideas about nonhuman or superhuman agents with extraordinary mental capacities (such as deities who are all-knowing) or agents with exceptional perceptual capacities (such as animals with specialized senses or superheroes with X-ray vision). How and when do such concepts of extraordinary or exceptional agents develop? Addressing this question promises to inform fundamental issues in cognitive development, such as the nature of intuitive and counterintuitive ideas, the enculturation of thought, and the cognitive foundations of religion.

Arguably, the ability to appreciate extraordinary or exceptional capacities originates in early childhood and is intimately linked to the development of an understanding of ordinary, human capacities. Children's understanding of persons and minds —their *theory of mind*—undergoes substantial development during the preschool years (Wellman & Liu, 2004), as children increasingly appreciate the subjective nature of perceptions and thoughts. Such development requires that children disambiguate how the world really is from how self and others perceive and think about the world. Very young children have difficulty understanding this distinction between appearance and reality. When asked about what other people know or believe, very young children tend to answer by simply assessing reality and using that information to infer others' knowledge and beliefs (Wellman, Cross, & Watson, 2001). We refer to this tendency as a *reality bias*.

One early-emerging manifestation of children overcoming this reality bias is their understanding of *ignorance*—an understanding that agents can be unaware of certain facts. By 3 years of age, children understand that ignorance may result from a variety of circumstances, including agents' lack of perceptual access to certain information. For example, older 3-year-olds understand that only those individuals who have looked inside a container will know what the container holds (Pillow, 1989; Pratt & Bryant, 1990). Soon after children develop an appreciation for the distinction between knowledge and ignorance, they begin to appreciate the distinction between reality and belief; they start to understand that others, misled by inaccurate perceptual cues or outdated information, can hold *false beliefs* (Wellman & Bartsch, 1988). In one standard false-belief task—an unexpected contents task (Perner, Leekam, & Wimmer, 1987)—children are shown that a familiar container (e.g., a cracker box) holds something atypical (e.g., rocks). The container is then closed and children are asked what another person, who has not seen in the container, will

---

Funding for this research was provided by National Institute of Child Health and Human Development Grant 22149 to Henry M. Wellman. We are very grateful to the children and parents who participated in this study and to the staff of the University of Michigan Children's Centers. We also thank two anonymous reviewers for their comments and suggestions.

Correspondence concerning this article should be addressed to Jonathan D. Lane, Department of Psychology, University of Michigan, 530 Church St., Ann Arbor, MI 48109. Electronic mail may be sent to jonlane@umich.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0012

think is inside. When faced with this task, 3-year-olds typically report that the other person will think the cracker box contains rocks, demonstrating a reality bias. Older 4-year-olds, in contrast, report that others will think the box contains crackers, demonstrating their ability to distinguish how the world really is from one's beliefs about the world.

Thus, in the preschool years, children evidence an emerging understanding of human knowledge and beliefs, including some appreciation of mechanisms that yield knowledge and beliefs (e.g., that seeing leads to knowing). But how does an understanding of the mental capacities of nonhuman or superhuman agents emerge? Studying children's developing understanding of extraordinary minds is interesting in its own right and also promises to shed light more generally on children's understanding of ordinary minds. One possibility, stemming from Piaget's (1929/1979) notion that preoperational children anthropomorphize all agents, has been termed the *anthropomorphism hypothesis*. According to this hypothesis (Boyer, 1996), when children first come to attribute constrained knowledge and fallible beliefs to ordinary humans, they attribute similar limitations to all agents. Only later, building on this initial platform, do children differentiate between the limited capacities of ordinary humans and extraordinary agents' less limited capacities.

An anthropomorphism hypothesis is intuitively appealing; even adults tend to think of nonhuman beings, such as God, as human-like (Barrett & Keil, 1996; Gray, Gray, & Wegner, 2007). Indeed, although formal religious doctrine may attribute radically nonhuman, counterintuitive capacities to deities (e.g., total omniscience), in everyday judgments adults tend to think of deities in terms that are more human-like and limited, and thus only moderately counterintuitive (Barrett, 2000; Boyer, 1996, 1998). For example, Barrett and Keil (1996) found that even religious believers well versed in God's omniscience conceptualized God as attending to people's prayers sequentially, rather than attending to all prayers simultaneously, suggesting that they thought of God as subject to some of the same spatiotemporal constraints as humans. If such anthropomorphic tendencies are true of adults, they seem all the more plausible for children.

However, seminal studies by Barrett, Richert, and Driesenga (2001) challenged this anthropomorphism perspective. These researchers asked children (who attended Christian preschools) to reason about the knowledge and beliefs of humans, God, animals, and inanimate objects. In two studies, using unexpected-contents false-belief tasks, 3- to 7-year-olds reported what their mother, a tree, nonhuman animals, and God would think is inside a cracker box that contained rocks. Regardless of age or level of false-belief understanding, children typically reported that God would know the actual contents of the box. In another study, these researchers asked 3- to 8-year-olds whether a human, a monkey, God, or a cat that could see in the dark would know the contents of a box that had only a small slit to peer inside and no internal illumination. Children consistently reported that God and the cat would know the contents of the box, both before and after they gained an understanding (at about age 5 years) that humans and monkeys would not know the contents of the box.

Based upon these findings, Barrett and colleagues (Barrett & Richert, 2003; Barrett et al., 2001) have endorsed an alternate *preparedness hypothesis*, which states that "early-developing conceptual structures in children used to reason about God are *not* specifically for representing humans, and, in fact, actually facilitate the acquisition and use of many features of God concepts of the Abrahamic monotheisms" (Barrett & Richert, 2003, p. 300). Further, Barrett et al. (2001) proposed that, "children can have a more accurate understanding of God's agency than that of humans" (p. 54). That is, very young children's tendency to *not* attribute false-beliefs or ignorance to any agent reflects an early supposition of infallible mental capacities; children treat all agents (human and nonhuman) as omniscient. This hypothesis thus advances the intriguing idea that early cognitive biases facilitate rapid awareness of certain counterintuitive ideas, including ideas about the extraordinary qualities of God.

Although intriguing, these findings and the preparedness hypothesis raise several questions, both empirical and conceptual. First, when very young children apparently attribute infallible knowledge and beliefs to persons (or Gods), their answers may simply reflect an early reality bias—they answer by reporting the reality of the situation without considering agents' mental abilities (Evans & Wellman, 2006; Wellman & Bartsch, 1988). The critical question then is what children attribute to God when they *first* start to distinguish between the actual state of reality and people's (often inaccurate) mental representations of that reality; in particular, at the point when they begin to attribute false beliefs or ignorance to humans. At that point, do they attribute fallible knowledge and beliefs to God as well, as proposed by the anthropomorphism hypothesis, or infallible knowledge and beliefs, as implied by the preparedness hypothesis? At a later age, when children have developed a more robust understanding of fallible mental capacities, around 5 to 6 years of age, both preparedness and anthropomorphism hypotheses might predict that children attribute more infallible mental capacities to God, provided they have been exposed to such information about God. The preparedness hypothesis predicts that such an understanding at age 6 would reflect a *continuation* of children's early default understanding of extraordinary minds. The anthropomorphism hypothesis, in contrast, posits that such an understanding at age 6 indicates that children are beginning to loosen their earlier tendencies to anthropomorphize all agents. To best test these two hypotheses, it is necessary to densely sample children within the proper age range. Further, the data should be analyzed in a sensitive age-related fashion in order to find and assess the critical window when children *first* correctly attribute fallible mental capacities to humans. Barrett et al. (2001) simply grouped 3-, 4-, and 5-year-olds in year-long age blocks that might have masked the critical developmental window during which children first begin to ascribe human-like limitations to nonhuman agents for these social-cognitive tasks.

A second empirical issue concerns the replicability of Barrett et al.'s (2001) findings. This may well be related to the first issue of fine-grained age sampling and analyses, because different samples or different age groupings may differentially capture the critical developmental window. On the one hand, Barrett and colleagues (Knight, Sousa, Barrett, & Atran, 2004) have replicated their findings with a sample of Yukatek Mayan children whose culture has adopted the Catholic God. As well, Richert and Barrett (2005) reported data conforming to a preparedness trajectory for children's performance on a diverse set of knowledge-ignorance tasks. But other researchers offer findings that conflict with those of Barrett and colleagues, and which would be better explained by the anthropomorphism hypothesis (e.g., Giménez-Dasí, Guerrero, & Harris, 2005; Makris & Pnevmatikos, 2007). For example, using a less challenging knowledge-ignorance task, Makris and Pnevmatikos (2007) found that 3- and 4-year-olds reliably attributed ignorance both to a human and to God. These latter findings suggest that, at least for certain mental properties, there may be a developmental period during which young children concurrently believe that human and nonhuman mental capacities are constrained.

Alongside these empirical issues is an important conceptual one. When children attribute accurate knowledge to God in a false-belief or knowledge-ignorance task, how are they reasoning about such knowledge and beliefs? One interpretation, following from the preparedness hypothesis, is that they are attributing (and prepared to attribute) to God something like omniscience—the capacity to know all things without perceptual access. But the tasks used tell us little about children's appreciation for the mechanisms through which agents acquire their knowledge or beliefs. In past studies, it is unclear whether young children (even 5-year-olds) attributed to extraordinary agents privileged knowledge directly or via certain (ordinary or exceptional) mechanisms. For example, children may have assumed that God had special visual capacities (a moderate extension of human capacities) and thus actually *saw* the contents of the containers. Indeed, Richert and Barrett (2005) found that children as young as 4 years understood that agents with specialized senses (e.g., exceptional hearing or vision) can gain knowledge about certain stimuli that would be elusive to normal humans. Thus, one unaddressed issue concerns when exactly children are able to understand that an agent can possess certain (privileged) knowledge or beliefs *without* the use of perceptual mechanisms. And this is related to whether and when they might attribute anything like omniscience to God.

We addressed these empirical and conceptual issues in several ways. We employed fine-grained age sampling and analyses (on the order of months rather than years) in an effort to reveal, more precisely, the ontogenetic unfolding of concepts of extraordinary minds. We tested children on both false-belief and knowledge-ignorance tasks and, because children develop an appreciation for human agents' ignorance and false beliefs on different timetables, we analyzed the age-related trajectories for these two tasks separately with appropriately different age groupings. Further, and crucially, we addressed the conceptual issue about children's understanding of mechanisms mediating ordinary and apparently extraordinary knowledge and beliefs. Our primary method was to present 3-, 4-, and 5-year-olds with carefully contrasting agents. Specifically, we asked children to report on the mental capacities of ordinary humans (their mother and a young girl), and various nonhuman and superhuman beings, including God. In some cases, we carefully specified, for the child, the capacities and mechanisms possessed by the agent. Two agents were specified as having special perceptual mechanisms for acquiring knowledge: a cat that can see in the dark and a superhero

(Heroman) who possesses X-ray vision. One other agent (Mr. Smart) was described as being able to "know everything," even without seeing, so Mr. Smart's knowledge or beliefs did not depend on perceptual mechanisms at all.

Each of these "special" agents was described in a brief but detailed way (see the Appendix). Mr. Smart's and Heroman's special powers were also elaborated through brief demonstrations. Note that Mr. Smart was of interest in his own right and also served as a control in relation to God. It is impossible to know the precise information each child had already received about God and surely children did not have equal exposure to tutelage about a sentient God or God's extraordinary attributes. Therefore, in contrast to God, about whom we provided no information, for Mr. Smart we gave all children exactly the same background information regarding his attributes. If, as predicted by the preparedness hypothesis, children are prepared to pick up on such information, given prior tendencies to think of agents as infallible and all-knowing, then Mr. Smart should be especially easy to appreciate. This set of contrasting agents allowed us to assess the extent to which children appreciated different mental and perceptual capacities as well as the specific mechanisms agents may use to gather information. As a second method for generating information about children's reasoning about mechanisms, we asked children to explain their judgments. For example, if they judged that an agent knew the contents of a completely darkened container, we asked them how that agent knew that information.

Method

Participants

Fifty-seven children (32 males), ranging in age from 40 to 73 months (*M* age = 54 months), participated. Primarily, children were of European American descent and lived in a middle- to upper-middle-class Midwestern university community. We densely sampled children in a critical age range (50–56 months) on the hypothesis (established during pilot testing) that, during this period, children would be likely to first understand ordinary humans' limited mental and perceptual capacities. One child was excluded from the sample because she could not remember the actual contents of the box for the knowledge-ignorance task. Following the interview, when asked what they knew about God, more than half of these children (59%) provided specific details about God (e.g., "He's very smart," "He's magical and powerful"). Thirty-nine parents agreed to briefly report on their child's exposure to religious concepts. Almost half of these parents reported that they take their child to a place of worship, most on a weekly or monthly basis.

Procedure

Children were interviewed individually in a quiet location. They were asked about the beliefs and knowledge of various agents (see the Appendix) using two tasks: one a contents false-belief task (Perner et al., 1987) and the other a knowledge-ignorance task (similar to that used by Barrett et al., 2001). Half of the children received the false-belief task first; the others received the knowledge-ignorance task first. Each of the agents was displayed on a 5 × 7 in. laminated card. Importantly, prior studies have shown that children perform equally well on these tasks whether the protagonists are drawings, puppets, or live performers (Wellman et al., 2001). Children were introduced to each agent (see precise language in the Appendix) upon their first exposure to that agent. For each task, children were presented either mom or the girl first. The presentation of the remaining agents was randomized (the girl or mom was presented as the second agent for 2 children, only).

Because prior research suggests that adults who are exposed to anthropomorphic images of God are more prone to make anthropomorphic judgments about God (Barrett & VanOrman, 1996), we assessed whether the presentation of an image representing God would influence children's judgments about God's mental capacities. Thus, half of the sample was tested using a blurry nondescript image to represent God (see the Appendix), and the other half received no image and no language alluding to a bodily presence for God.

Measures

False-belief understanding. Children were shown a crayon box and a brown paper bag. The experimenter asked children what they thought was inside the crayon box and then showed them that the box actually held marbles and that the paper bag held crayons. Both containers were closed, and as a memory check, children were asked which container had marbles and which container had crayons inside (all children answered correctly). Children were then asked the following for each agent with regard to the box: "__ has never been in the room with these things before. If we show __ this box, all closed up, [picture of agent approaches the crayon box] what will __ think is inside here?'' To minimize anthropomorphic cues, half of the children were not shown a picture representing God nor were told that God had ''never been in the room''; they were simply asked, ''What will God think is inside here?'' To deemphasize Mr. Smart's visual abilities and emphasize his all-knowing capacity, children were asked what Mr. Smart would think is in the box if he stayed across the room, facing away from the box. For each agent, children earned a score of 0 if they attributed a correct belief or 1 if they attributed a false belief. Following each judgment, the interviewer prompted children to justify their answer by asking, ''Why will __ think __ are inside?''

Knowledge-ignorance understanding. To assess children's ability to distinguish knowledgeable versus ignorant agents, children were shown two boxes, each with a slit at the top allowing children to look inside. A lamp was positioned above each box. One of the lamps was turned on, illuminating the interior of the corresponding box and revealing a red plastic frog inside. The other lamp was off, and the corresponding box appeared completely empty. Children were first asked to look inside the lit box and to report what they saw. After children reported that the lit box contained a red frog, they were asked to look inside the unlit box. After children reported that they could not see anything inside the unlit box, the experimenter turned on the corresponding lamp, revealing an identical red frog inside. The latter light was then turned off, and the experimenter reminded children, ''So, both boxes have a frog inside but you can't see the frog when this one is dark'' [pointing at the unlit box]. As a memory check, children were asked whether each box, in turn, contained a frog or was empty (corrective feedback was provided for a few children as necessary). Children were then asked the following for each agent with regard to the unlit box: ''__ has never been in this room with these boxes before. If __ comes very close to the top of the dark box, what will __ think is inside here; a frog or nothing?'' For this focal question, the experimenter held the agent's picture above the unlit box, facing the contents of the box. To minimize anthropomorphic cues for half of the sample, no picture representing God was presented and children were not told that God had ''never been in this room''; they were simply asked, ''What will God think is inside here; a frog or nothing?'' while the interviewer pointed at the box. To deemphasize Mr. Smart's visual abilities, children were asked what Mr. Smart would think is in the box if he stayed across the room, facing away from the box. For each agent, children earned a score of 0 if they attributed correct knowledge or 1 if they attributed ignorance. Following each judgment, the interviewer prompted children to justify their choice by asking, ''Why will __ think a frog is inside?'' or ''Why will __ think nothing is inside?''

Results

Before conducting our focal analyses, we determined whether the presentation of an image representing God (along with language alluding to God's physical presence) was related to children's attribution of human-like (i.e., fallible or constrained) mental capacities to God. For the false-belief task, 43% of those children presented the image and language, and 36% of those who neither saw the image nor heard the language, reported that God will think crayons are in the crayon box, $\chi^2(1, 56) = 0.30$, *ns*. For the knowledge-ignorance task, 61% of children presented the image and language, and 64% of those not presented the image or language, reported that God will think nothing is inside the unlit box, $\chi^2(1, 56) = 0.08$, *ns*. Because children who were presented the ''anthropomorphic'' image and language and those not presented the image and language were equally likely to ascribe fallible capacities to God, they were combined in the focal analyses.

False-Belief Understanding

To be clear in what follows, a ''correct belief'' refers to the judgment that an agent knows what is actually in the box (its real albeit hidden contents—marbles). ''False beliefs'' refer to judgments that an agent is mistaken (i.e., will think crayons are inside the crayon box). Preliminary analyses assessed whether some children concurrently attributed human-like, false beliefs to normal humans *and* to extraordinary beings, as expected under an anthropomorphism hypothesis. Many children indeed attributed false beliefs to each of the special agents—Heroman, Mr. Smart, and God—as well as to normal humans (i.e., the girl and mom) at levels significantly different from chance (for details, see Table 1). Of primary interest was whether a pattern of attributing human-like limitations to all agents would be most common when children first began to understand that humans may hold false beliefs. We conducted an

Table 1
Chi-Square Analyses Comparing Children's Attribution of False Beliefs to Both Agents Against Chance

Agent	Heroman	Mr. Smart	God
Mother	$\chi^2(1, 18) = 8.00^{**}$	$\chi^2(1, 21) = 10.71^{**}$	$\chi^2(1, 22) = 14.73^{***}$
Girl	$\chi^2(1, 18) = 5.56^{*}$	$\chi^2(1, 21) = 8.05^{**}$	$\chi^2(1, 22) = 11.64^{***}$

Note. Eighteen children attributed a false belief to Heroman, 21 to Mr. Smart, and 22 to God.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

exploratory analysis of children's false-belief judgments to find an age range during which children typically attributed "correct" beliefs to all agents and an immediately subsequent period during which children attributed false beliefs to ordinary humans. Based upon this exploratory analysis, we divided children into three age groups: 24 in the young group (40.4–52.4 months; *M* = 47.7), 17 in the middle group (52.5–58.9 months; *M* = 54.7), and 15 in the oldest group (59.0–73.4 months; *M* = 63.3). Figure 1 depicts the primary data: the percentage of children who attributed a false belief to each agent, by age group. Three trends are apparent in this graph: (a) with increasing age, children more often attributed false beliefs to mom and the girl; (b) children in all three age groups ascribed correct beliefs to Heroman; and (c) the youngest and oldest children attributed correct beliefs to God and Mr. Smart, whereas children in the middle age group typically attributed false beliefs to God and Mr. Smart.

Judgments. An initial repeated-measures analysis of variance (ANOVA) for children's attributions of false beliefs with age as a between-subjects factor (3: young, middle, old), and agent as a within-subjects factor (5: mom, girl, Mr. Smart, Heroman, God) revealed a significant effect for age, $F(2, 53) = 6.98$, $p < .01$, and agent $F(4, 212) = 10.50$, $p < .001$, and a significant interaction between age and agent, $F(8, 212) = 5.13$, $p < .001$. For the youngest group, there were no differences between the agents in children's attributions of false beliefs, $F(4, 92) = .57$, *ns;* children reported that all agents would think the crayon box contains marbles. On a composite measure summing responses for both the ordinary, human agents, these youngest children attributed "correct" beliefs at levels significantly above chance, $t(23) = 2.70$, $p < .05$, and did so as well on a similar measure summing responses for both Mr. Smart and God, $t(23) = 3.41$, $p < .01$. Children in the middle group attributed false beliefs to each agent (except Heroman) more often than did the youngest children, $ts(39) > 2.70$, $ps < .05$. In this group, children affirmed false beliefs at levels above chance on a composite measure of judgments for both ordinary, human agents, $t(16) = 2.70$, $p < .05$. On a parallel measure, their attributions of false beliefs for God and Mr. Smart were similar to their judgments for mom and the girl, but not significantly above chance. Note, however, that

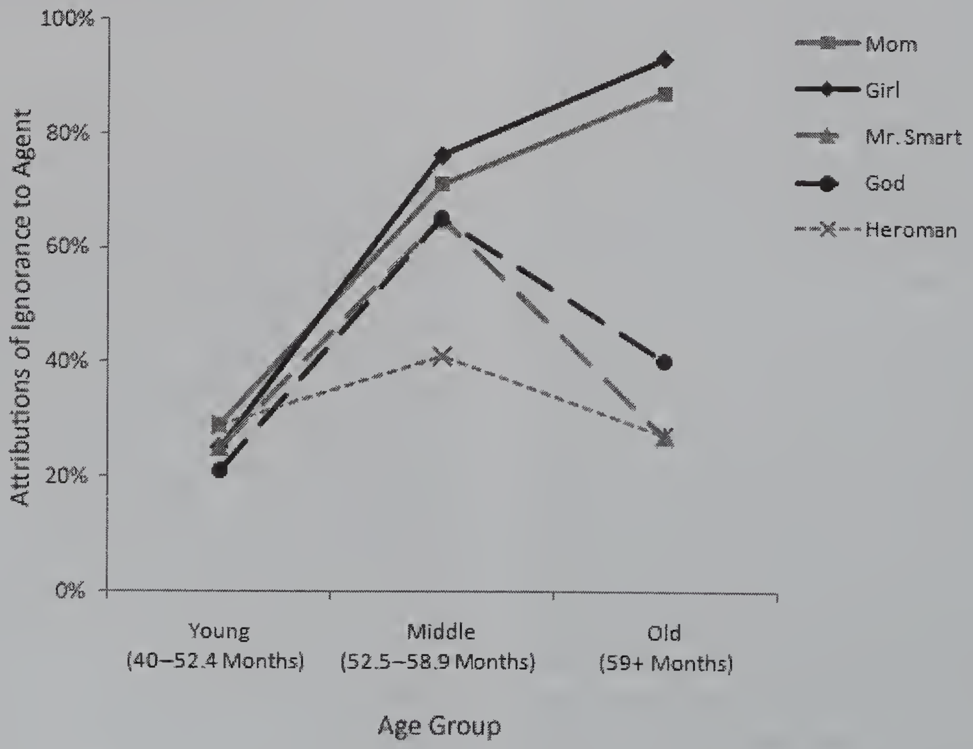


Figure 1. Percentage of children, by age group, reporting that the protagonist will hold a false belief (i.e., will think there are crayons in the crayon box).

Table 2
Coding of Open-Ended Responses

Category	Description	Examples: False-belief task	Examples: Knowledge-ignorance task
Reality based	Child cites actual contents of the containers. Child does not mention agent's mental or perceptual capacities	“There are marbles inside the box” “The crayons are in the bag” “I saw it in there”	“There is a frog inside” “I saw it in there”
Appearance based	Child cites the appearance of the box	“There are crayons on the box” “It looks like a crayon box” “It says 'crayons'”	“It's dark” “It looks like its empty”
Typicality based	Child cites the type of container or mentions what that the type of box typically holds	“It's a crayon box” “There's usually crayons in a crayon box”	“It's supposed to be in there” “Frogs live in boxes”
Exceptional perceptual capacities	Child cites agent's exceptional senses (vision, hearing) when justifying why the agent will think the true contents are inside	“He has special/x-ray/lit-up eyes” “He can see in the dark” “He heard us talking about the marbles”	“He can see in the dark” “Because he saw it” “Heard us talking about the frog”
Inadequate perceptual capacities	Child cites agent's inadequate, senses (e.g., inadequate vision)	“He can't see through things” “He can't see” “Can't see in closed things”	“It's too dark to see inside” “He can't see” “Light isn't on” [implies agent can't see]
Extraordinary mental capacities	Child cites agent's mental capacities without referring to perceptual capacities	“He's super smart” “He knows everything”	“He's very smart” “He thinks very good”
Inadequate mental capacities	Child cites agent's inadequate mental capacities without referring to perceptual capacities	“He doesn't know crayons are in there” “He doesn't know”	“He's not as smart as Smarty” “She won't know there's a frog inside'

children in this middle group *did not* grant God or Mr. Smart correct beliefs—contrary to what would be expected by a preparedness hypothesis. Only the oldest children (those 59 months and older) consistently reported that Mr. Smart and God would possess correct beliefs whereas the beliefs of mom and the girl would be false, ts(14) > 3.50, $p < .01$. Children in the middle group, as in the oldest group, affirmed correct beliefs for Heroman (tending to report that Heroman will know the crayon box actually contains marbles) in contrast to the false beliefs of mom, $t(16) > 2.06$, $p < .06$, and the girl, $t(16) > 2.40$, $p < .05$.

Justifications. Children's justifications help clarify the reasoning behind their judgments. Children's justifications were coded into seven categories, as outlined in Table 2, and a residual uncodable category (“don't know”). Twenty percent of the justifications were coded by two separate coders (one blind to all hypotheses and aims of the study) to assess interrater reliability (all κs ≥ .96). We focus on three contrasting agents: Heroman (whose special vision was described), Mr. Smart (who was described as having an extraordinary mind, but no specific exceptional perceptual mechanism), and God (about whom we told children nothing). Table 3 presents data on the primary forms of reasoning that children used to justify these three agents' correct beliefs. Justifications for correct beliefs are particularly revealing because they address which (if any) extraordinary capacities children attributed to these agents.

The youngest children generally did not appreciate these agents' mental capacities or knowledge-collecting mechanisms; rather, they evidenced a reality bias (or said something uninformative). Most of the youngest children said that Heroman would know that the box contains marbles, but fewer than a third referred to his exceptional vision (e.g., “He has super eyes”). Rather, most of these

Table 3
Justifications for Attributions of Correct Beliefs and Knowledge

		Correct belief reasoning				Correct knowledge reasoning			
Age group	Agent	Reality based	Exceptional perceptual capacities	Extraordin. mental capacities	Uninformative	Reality based	Exceptional perceptual capacities	Extraordin. mental capacities	Uninformative
Young	Heroman	24%	29%	6%	35%	29%	14%	14%	43%
	Mr. Smart	17%	6%	28%	33%	13%	—	13%	50%
	God	32%	5%	5%	37%	17%	—	17%	50%
Middle	Heroman	10%	60%	10%	—	18%	73%	—	—
	Mr. Smart	17%	—	67%	17%	33%	—	67%	—
	God	50%	17%	17%	—	67%	—	—	33%
Oldest	Heroman	—	100%	—	—	33%	67%	—	—
	Mr. Smart	—	—	100%	—	35%	5%	60%	—
	God	—	22%	56%	—	42%	25%	17%	17%

Note. Age ranges for the false-belief task are: young (40.4–52.4 months), middle (52.5–58.9 months), and oldest (59.0–73.4 months). Age ranges for the knowledge-ignorance task are: young (40.4–49.4 months), middle (49.5–54.5 months), and oldest (54.6–73.4 months). Extraordin. mental capacities = extraordinary mental capacities.

children justified Heroman's correct beliefs by citing reality (e.g., "There is crayons inside"), confabulating a reason (e.g., "He guessed"), or provided an uninformative response (e.g., "Because" or "I don't know"). In this young group, of the children who said that Mr. Smart or God would think that the crayon box contains marbles (correct belief), only a few provided justifications indicating an appreciation for Mr. Smart's or God's extraordinary mental capacities (e.g., "He's super smart"). Of the remaining children who attributed correct beliefs to Mr. Smart and God, most justified these agents' correct beliefs by citing reality or provided an uninformative response. In sum, when attributing correct beliefs to these special agents (as was typical for these young children, who generally attributed correct beliefs to *all* agents), only 22% of children's responses mentioned anything like exceptional perception or extraordinary mental capacities, although such special capacities were described to children at several points in our protocol.

In the middle group, most children attributed a correct belief to Heroman, and more than half justified their answers by specifically referring to Heroman's extraordinary vision (e.g., "He can look through things"). However, most children attributed to all other agents a false belief. Of the 11 children who attributed a false belief to Mr. Smart, 10 (91%) cited the appearance of the box (e.g., "There's a picture of crayons on the box") or the type of box (e.g., "It's a crayon box"). Similarly, of the 11 children who attributed a false belief to God, 9 (82%) cited the appearance of the box or the type of box. When attributing correct beliefs to these special agents, children in this middle group cited exceptional perception or extraordinary mental capacities in 55% of their responses.

Just as in their judgments, the oldest children's justifications suggested an appreciation for Heroman's perceptual prowess *and* for Mr. Smart's and God's extraordinary minds. All children in the oldest group who attributed a correct belief to Heroman cited his extraordinary vision, and every child who attributed a correct belief to Mr. Smart cited his mental capacities (e.g., "He's so smart"). Interestingly, of the children who attributed a correct belief to God, about half cited extraordinary mental capacities (e.g., "He knows everything"), but several attributed to God exceptional visual capacities (e.g., "He can see through anything"). Overall, when attributing correct beliefs to these special agents (the typical response for the oldest children), older children appealed to exceptional perception or extraordinary mental capacities 94% of the time. In sum, when attributing correct ("infallible") beliefs to these agents, young children rarely referred to agents' extraordinary perceptual or mental capacities; however, this sort of reasoning was provided often by the middle age group and especially by the oldest children.

Knowledge-Ignorance Understanding

With regard to children's understanding of agents' ignorance, we again first determined whether some children concurrently attributed human-like, constrained capacities to most agents. For clarification, we use the phrase ''correct knowledge'' when referring to children's judgments that the agent knows what is actually in the box (its real but hidden contents—a frog). ''Ignorance'' refers to children's judgments that the agent is mistaken (i.e., will think the box is empty). A significant number of children attributed ignorance to each of the special agents—Heroman, the cat, Mr. Smart, and God—as well as to the ordinary humans (i.e., the girl and mom), at levels significantly different from chance (see Table 4 for details). Again, of primary interest was whether this pattern would be most common among children who were just beginning to understand that ordinary humans' knowledge can be limited by perceptual access. We conducted an exploratory analysis of children's knowledge-ignorance judgments to find an age range when children typically attributed ''correct'' knowledge to all agents, and a subsequent period when children attributed ignorance to ordinary humans. Based on these preliminary analyses, we divided children into three age groups. Because children evidenced an understanding of ignorance at an average age of 4 months younger than an understanding of false beliefs, for knowledge-ignorance analyses there were 12 children in the young group (40.4–49.4 months; *M* = 44.2), 20 in the middle group (49.5–54.5 months; *M* = 52.2), and 24 in the oldest group (54.6–73.4 months; *M* = 60.5). Figure 2 depicts the focal data: the percentage of children who attributed ignorance to each agent, per age group. Three trends are noticeable in this graph: (a) with increasing age, children more often attributed ignorance to mom and the girl, with a later leveling off; (b) children in all three age groups attributed correct knowledge to Heroman

Table 4
Chi-Square Analyses Comparing Children's Attribution of Ignorance to Both Agents Against Chance

Agent	Cat	Heroman	Mr. Smart	God
Mother	$\chi^2(1, 14) = 4.57^*$	$\chi^2(1, 20) = 7.20^{**}$	$\chi^2(1, 22) = 6.55^*$	$\chi^2(1, 35) = 17.86^{***}$
Girl	$\chi^2(1, 14) = 7.14^{**}$	$\chi^2(1, 20) = 12.80^{***}$	$\chi^2(1, 22) = 18.18^{***}$	$\chi^2(1, 35) = 27.46^{***}$

Note. Fourteen children attributed ignorance to the cat, 20 to Heroman, 22 to Mr. Smart, and 35 to God.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

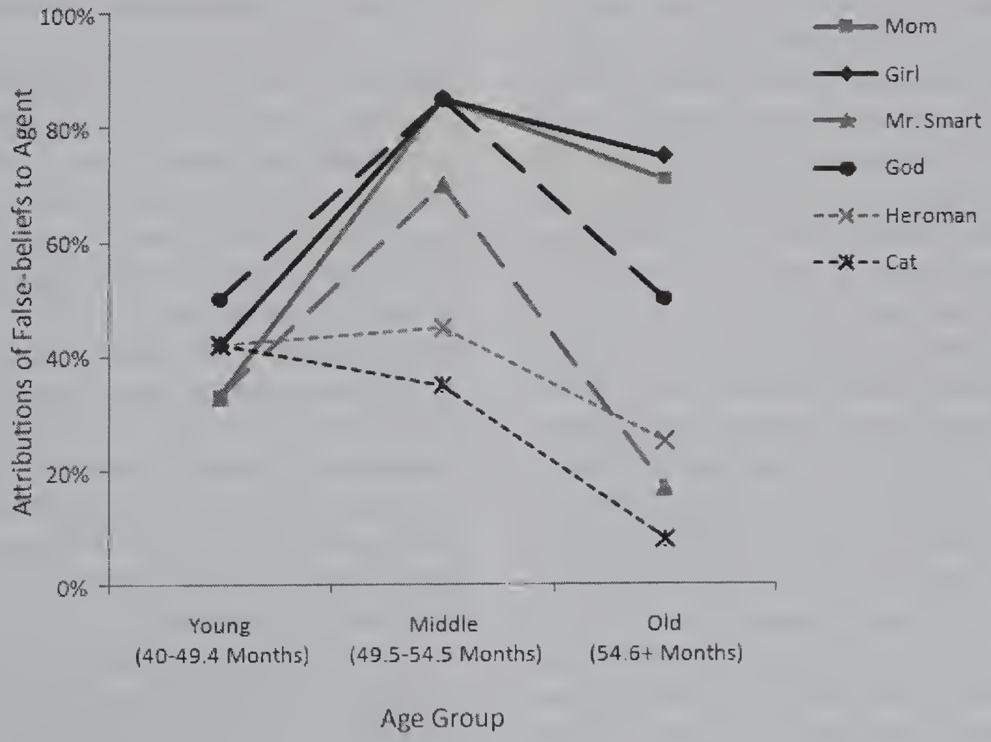


Figure 2. Percentage of children, by age group, reporting that the protagonist will not perceive the contents of the unlit box (i.e., will think that the unlit box is empty).

and the cat, a tendency that was particularly pronounced in the oldest age group; and (c) children in the young and old age groups attributed correct knowledge to God and Mr. Smart, whereas children in the middle age group attributed ignorance to God and Mr. Smart.

Judgments. A repeated measures ANOVA for children's attributions of ignorance with age as a between-subjects factor (3: young, middle, old) and agent as a within-subjects factor (6: mom, girl, Mr. Smart, Heroman, God, cat), revealed a significant effect for age, $F(2, 53) = 4.73$, $p < .05$; agent, $F(5, 265) = 11.16$, $p < .001$; and an interaction between age and agent, $F(10, 265) = 3.58$, $p < .001$. For the youngest children, there were no differences in children's attribution of ignorance between the agents, $F(5, 55) = .61$, *ns*, and they attributed ignorance to all agents at chance levels. Children in the middle group attributed ignorance to agents (except Heroman and the cat) more often than did the youngest children, $ts(30) > 2.09$, $ps < .05$. They did so above chance on a composite measure of judgments for both ordinary, human agents, $t(19) = 6.66$, $p < .001$, and notably were above chance on a similar composite measure for both Mr. Smart and God, $t(19) = 3.24$, $p < .01$. These children (along with the oldest children) distinguished between the correct knowledge of Heroman and the cat and the ignorance of mom and the girl, $ts(19) > 2.99$, $ps < .01$. Notably, in the middle age group, only the two agents who were specified as possessing exceptional vision were judged to know the correct contents of the dark box; God and Mr. Smart were judged to be ignorant. Only the oldest group of children (54 months and older) consistently reported that whereas mom and the girl would think the unlit box is empty, Mr. Smart and God would have correct knowledge that the unlit box contains a frog, all $ts(23) > 2.77$, $ps \leq .01$, except for the difference between mom and God, which was marginally significant at $p < .06$.

Justifications. Children's justifications shed light on the reasoning behind their judgments. Justifications were coded into the same seven focal categories as before (see Table 2). Twenty percent of the justifications were coded by two coders (one unaware of the hypotheses and aims of the study) to assess interrater reliability (all $\kappa s \geq .88$). We focus on four contrasting agents: Heroman and the cat (whose exceptional visual abilities were described), Mr. Smart (whose extraordinary mind was described but who was not given a special perceptual mechanism), and God (about whom we told children nothing). Table 3 presents data on the most common forms of reasoning that children used to justify Heroman's, Mr. Smart's, and God's correct knowledge.

For the youngest children, most judged that Heroman and the cat would know that the box contains a frog, but only 1 of these children for each agent referred to special vision ("He can use his light sensy" and "He can see in the dark"). In contrast, most of these children justified these agents' correct knowledge by citing reality (e.g., "I saw a frog") or provided an uninformative response (e.g., "Because" or "I don't know"). In this young group, of the many children who reported that Mr. Smart would know that the unlit box contains a frog, only 1 provided a justification that suggested he appreciated Mr. Smart's extraordinary mind ("He knows everything"), and only 1 of the children who reported that God would know that the box contains a frog cited God's mental capacities ("He knows everything at church"). The remaining children justified Mr. Smart's and God's correct knowledge by citing the real nature of the box, or provided uninformative responses. In summary, when attributing correct knowledge to these special agents (the typical response for the young children, who generally attributed correct knowledge to *all* agents), in only 21% of their responses did the youngest children appeal to exceptional perception or extraordinary mental capacities.

In the middle group, most judged Heroman and the cat to have correct knowledge. Almost all these children provided justifications that specifically referred to these agents' exceptional vision (e.g., "He can see through anything" and "He can see in the dark," respectively). Most children attributed to all other agents ignorance. Of the 14 children who attributed ignorance to Mr. Smart, 4 (29%) justified their responses by citing the appearance of the box (e.g., "It's dark") and 7 (50%) referred to Mr. Smart's inadequate visual capacities (e.g., "It's hard to see"). Similarly, of the 17 children who attributed ignorance to God, 3 (18%) referred to the appearance of the box (e.g., "It's dark") and 10 (59%) justified their responses by referring to God's inadequate visual capacities (e.g., "He can't see in the dark"). When attributing correct beliefs to these four special agents, children referred to exceptional visual capacities or extraordinary mental abilities in 73% of their responses.

The oldest children's justifications suggested a greater appreciation for Mr. Smart's and God's extraordinary mental capacities, in addition to an appreciation for Heroman's and the cat's exceptional vision. Most of the oldest children judged Heroman and the cat to have correct knowledge, and a majority of these children justified their judgments by explicitly referring to these agents' exceptional visual capacities. Of the children in this group who attributed correct knowledge to Mr. Smart, more than half cited his extraordinary mental capacities (e.g., ''He knows everything''), 1 referred to exceptional visual capacities (''He can see anything''), and a third cited reality. Of the children who attributed correct knowledge to God, almost half cited reality, whereas others cited extraordinary mental capacities (e.g., ''He knows everything'') or exceptional visual capacities. In short, when attributing correct knowledge to these four agents (the typical response for the oldest children), the oldest children cited exceptional perception or extraordinary mental abilities in 63% of their responses.

Discussion

Not only do children everywhere come to understand the basic mental and perceptual capacities of ordinary human agents, they come to entertain ideas about agents with extraordinary capacities, including religious deities. Recent research on children's understanding of extraordinary, nonhuman minds has been aimed at comparing and contrasting two opposing positions: the anthropomorphism hypothesis and the preparedness hypothesis. From the *anthropomorphism* perspective, very young children initially understand all intentional agents as possessing human-like capacities. Three-year-olds fall prey to a reality bias—they do not consider agents' mental abilities and thus fail to distinguish between the state of reality and people's (often inaccurate) mental states. As children begin to appreciate that humans' capacities are limited and fallible, they initially attribute similar limitations to all agents. After developing an understanding of ordinary, human capacities, children can appreciate that certain agents may conceivably have exceptional or extraordinary powers. Thus, to understand extraordinary agents' special capacities, children must overcome or modify their intuitive conceptions of agents and increasingly think in counterintuitive, nonanthropomorphic terms. The alternative *preparedness* perspective proposes that very young children begin life well equipped to understand that certain agents possess nonhuman capacities (e.g., infallible beliefs) because they have an initial ''default assumption'' that ''many superhuman properties are the norm'' (Richert & Barrett, 2005, p. 284). Because initially all beliefs are true and all agents infallible, for special agents (God or others when they are explicitly told the agent has superhuman capacities or states) children need merely and easily continue to see them as superhuman.

A crucial developmental difference exists for these two positions. According to the anthropomorphism hypothesis, there should exist a developmental point, once children begin to attribute fallible capacities and states to humans, when they attribute these states to all agents, even agents adults contend are infallible (e.g., God) and children have heard are infallible (perhaps God, but in our tasks explicitly Mr. Smart). For the preparedness hypothesis, on the other hand, because superhuman properties are the default ''norm,'' children need not struggle to reason counterintuitively about such agents and, as a consequence, there should be no developmental point when these agents are attributed fallible, limited capacities or states. For both positions, an intriguing contrast case concerns agents with more limited extension of ordinary capacities, such as night vision or X-ray vision.

In accordance with the anthropomorphism hypothesis, we found that children reliably attributed ordinary, human-like capacities to special agents for both a false-belief task and a knowledge-ignorance task. Children not only did this for God but also did so for agents whose extraordinary mental capacities and exceptional perceptual mechanisms were explicitly described and demonstrated to them. Normatively, children come to understand ignorance before understanding false beliefs; this is true in precise scaling comparisons (e.g., Wellman & Liu, 2004) and is apparent in our own data as well. Accordingly, we found that the age-related developmental period during which children were especially likely to consider most agents' capacities to be human-like differed between these two forms of mental understanding, a sequential pattern that seems to accord more with an anthropomorphism position. A prepared, early understanding of infallibility should generally apply to knowledge and belief; overcoming an early reality bias could more sensibly apply first to developmentally ''easier'' mental states (ignorance), then more complicated mental states (false beliefs). Thus, it is of note that our data show that these children came to understand the ordinary limits of one and then another mental capacity, and subsequently entertained the counterintuitive suspension of those limits in sequence.

In apparently contradicting the results and conclusions of Barrett and colleagues, the current findings join two other recent studies. Using a different type of knowledge-ignorance task with Greek Orthodox children, Makris and Pnevmatikos (2007) found that 3- and 4-year-olds held that God and a little girl would *both not know* the contents of a closed box. Only at age 5 did participants differentiate between the girl's ignorance and God's correct knowledge of the box's true contents. Similarly, using modified knowledge-ignorance and false-belief tasks with a sample of Spanish children, Giménez-Dasí et al. (2005) found that, compared with 3- and 5-year-olds, their 4-year-olds more often attributed ignorance and false beliefs to God, and this was the case for children who were attending religious as well as those attending nonreligious preschools.

Our data go beyond other results, however, in clarifying how the overall pattern of apparently contradictory findings could arise depending on the way in which children were sampled and grouped. That is, our finding of specific ''anthropomorphic'' developmental periods (using tasks similar to those used by Barrett et al., 2001) was a product of careful age sampling and precise age-related analyses of the data. Suppose instead we reexamine our data after removing participants to correct for our dense sampling in the middle age range and simply divide our remaining children into three arbitrary age groups. Then only two age-graded linear trends emerge: (a) an increasing tendency to attribute fallible capacities to the girl and mom and (b) a constant trend to attribute ''infallible'' capacities to God, Mr. Smart, and Heroman. That is, with less precise sampling and age grouping, our data would mimic the trend lines shown in Barrett's work (e.g., Barrett et al., 2001; Knight et al., 2004). But as is clear in Figures 1 and 2, such analyses would actually mask three different trends: (a) an increasing linear trend to attribute fallible mental states to the girl and mom, (b) an apparently constant trend to attribute correct mental states to the cat and Heroman, and (c) a curvilinear trend where, with age, children first increasingly attribute fallible mental states to Mr. Smart and God, and only later attribute extraordinary, correct mental states to those superhuman agents. The latter trend thus parallels the findings of Makris and Pnevmatikos (2007), and also of Giménez-Dasí et al. (2005), and suggest that their data, too, may have emerged from samples and groupings that were able, like our own, to reveal more precise and detailed developmental trends.

Crucially, earlier studies provide limited information on what specific capacities children are attributing to extraordinary agents. The exception is Giménez-Dasí et al.'s (2005) finding that, when asked, 4- and 5-year-olds often justified their answers that God's knowledge was limited by referencing God's limited perceptual capacities. In our more comprehensive assessment of children's reasoning, we distinguished children's appreciation for mental capacities (e.g., infallible beliefs) from their appreciation for perceptual capacities (e.g., night vision). We used two techniques to provide this additional, needed, information: (a) asking children about the knowledge and beliefs of agents with contrasting mental and perceptual capacities and (b) asking children to explain their responses. These techniques jointly revealed that children in the middle age groups appreciated that most humans (such as the girl, mom, and Mr. Smart) as well as God have constrained access to certain information. However, they also reasoned that other agents who possess specific perceptual mechanisms to access information (such as the exceptional visual capacities of the cat and Heroman) may gain knowledge that would be elusive to most humans. It is conceivable that children attributed accurate knowledge and beliefs to Heroman merely because he held ''superhero'' status, and thus they understood him as being *all-capable* within a pretend world. However, on the knowledge-ignorance task, children treated Heroman just as they treated the cat—an agent who was not presented as a ''super'' agent. Moreover, for both tasks, children who attributed accurate knowledge and beliefs to Heroman justified their judgments by specifically referring to his *visual* abilities and not to his other exceptional abilities, such as his ability to fly fast.

Only the oldest children appreciated Mr. Smart's and God's extraordinary mental capacities in the absence of exceptional perceptual mechanisms through which knowledge or beliefs could be acquired—Mr. Smart was described as being very smart and knowing everything, but we did not tell children how he acquired information other than saying that it was not based on vision. These data thus provide converging evidence that contradict the preparedness hypothesis, which specifically describes superhuman capacities, such as infallible beliefs, as the default. According to the preparedness hypothesis, children should have simply continued with this default position when taught about Mr. Smart, but they did not. Similarly, only the oldest children understood the extraordinary mind of God, about whom we provided no background information. Perhaps, in comparison with religiously trained children, the children in our sample lacked knowledge of God's powers, and this is the reason why children in the middle age group tended to anthropomorphize God—they simply had not learned about God's powers. Following the interview, we did ask children what they know about God. Exploratory analyses did not indicate that there were differences in the responses of children based upon their knowledge of God. The children who could provide details about God, for example, were evenly distributed between the three age groups for the knowledge-ignorance task—about a third (35%) were in the middle ''anthropomorphic'' age group—and these children comprised about half (45%) of the children in the middle age group. Moreover, we found the same trends for God as we did for Mr. Smart, an agent about whom children definitely had relevant background information. However, our measures did not sensitively capture details of children's prior religious instruction, and our parental reports of children's religious exposure were available for only 39 of 56 children. It is important that future research be conducted using similar methods to assess these developing concepts in children who are exposed heavily to theistic ideas.

Although the current results go a long way toward clarifying children's developing ideas about extraordinary minds and exceptional perceptual capacities, they also raise several questions that could be addressed with future research. First, although we found that a substantial portion of children attributed fallible capacities to the two agents described as having exceptional perceptual capacities—Heroman and the cat—we did not find a specific developmental period during which children were particularly likely to anthropomorphize these agents. It is possible that an even finer-grained age sampling and analysis might uncover earlier anthropomorphic windows for children's understanding of exceptional perceptual capacities. Second, our findings do *not* indicate that our oldest children were attributing *omniscience* to Mr. Smart and God, although some researchers have used that term when interpreting their data (e.g., Giménez-Dasí et al., 2005; Knight et al., 2004). Understanding omniscience requires that one appreciates that an agent not only knows the contents of closed containers but also truly knows *everything*—the nature of all past and future events, all scientific facts, everyone's unspoken intentions and dreams, and so on. Rather, our oldest children likely understood God's and Mr. Smart's powers in a much less counterintuitive manner—Mr. Smart and God think like humans, but they know more (about certain tangible things in the here and now) than normal humans. It would certainly be interesting in future studies to address how children come to entertain increasingly counterintuitive concepts of extraordinary mental capacities and how this might eventually lead to a sophisticated understanding of omniscience.

In summary, our data lend critical support to the anthropomorphism hypothesis while also suggesting ways in which this hypothesis might be modified. When these children first began to overcome their initial reality bias and started to understand the limits of human mental capacities (e.g., that ordinary humans possess ignorance and false beliefs), they applied this same understanding to God and to an agent who was described as possessing extraordinary mental capacities. Meanwhile, during this ''anthropomorphic'' period, children appreciated that knowledge and beliefs may be acquired via highly specialized, nonhuman perceptual capacities—such as eyes that see in the dark. This initial grasp of exceptional perceptual capacities may well act as a bridge to a later understanding of even more counterintuitive, superhuman capacities. By age 5, children understood that agents with special mental capacities, such as the ability to know everything, may possess knowledge and beliefs without necessarily relying on specialized perceptual mechanisms for acquiring that information. Such early abilities to understand extraordinary capacities are indeed impressive. These abilities allow children to begin to grasp religious teachings that are seemingly counterintuitive, such as God's omniscience. But, importantly, this early understanding of the extraordinary is built upon an earlier and more fundamental understanding of the ordinary.

References

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*, 29–34.

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology*, *31*, 219–247.

Barrett, J. L., & Richert, R. A. (2003). Anthropomorphism or preparedness? Exploring children's god concepts. *Review of Religious Research*, *44*, 300–312.

Barrett, J. L., Richert, R. A., & Driesenga, A. (2001). God's beliefs versus mother's: The development of nonhuman agent concepts. *Child Development*, *72*, 50–65.

Barrett, J. L., & VanOrman, B. (1996). The effects of image-use in worship on God concepts. *Journal of Psychology and Christianity*, *15*, 38–45.

Boyer, P. (1996). What makes anthropomorphism natural: Intuitive ontology and cultural representations. *Journal of the Royal Anthropological Institute*, *2*, 83–97.

Boyer, P. (1998). Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist*, *100*, 876–889.

Evans, E. M., & Wellman, H. M. (2006). A case of stunted development? Existential reasoning is contingent on a developing theory of mind. *Behavioral and Brain Sciences*, *29*, 471–472.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 851–898). Hoboken, NJ: Wiley.

Giménez-Dasí, M., Guerrero, S., & Harris, P. L. (2005). Intimations of immortality and omniscience in early childhood. *European Journal of Developmental Psychology*, *2*, 285–297.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*, 619.

Knight, N., Sousa, P., Barrett, J. L., & Atran, S. (2004). Children's attribution of beliefs to humans and God: Cross-cultural evidence. *Cognitive Science*, *28*, 117–126.

Makris, N., & Pnevmatikos, D. (2007). Children's understanding of human and super-natural minds. *Cognitive Development*, *22*, 365–375.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year olds' difficulty understanding false-belief: Representational limitation, lack of knowledge or pragmatic understanding? *British Journal of Developmental Psychology*, *5*, 125–137.

Piaget, J. (1979). *The child's conception of the world*. Totowa, NJ: Littlefield, Adams. (Original work published 1929)

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*, 116–129.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*, 973–982.

Richert, R. A., & Barrett, J. L. (2005). Do you see what I see? Young children's assumptions about God's perceptual abilities. *International Journal for the Psychology of Religion*, *15*, 283–295.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

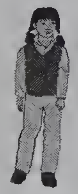
Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*, 239–277.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false-belief. *Child Development*, *72*, 655–684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541.

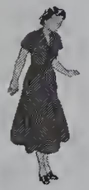
Appendix

Agent Images and Introductions



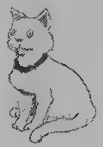
Girl

Let's talk about Mary. [Show picture of Mary]



Mom

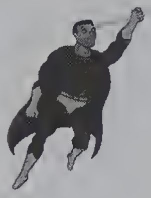
Let's talk about your mom. [Show picture of mom] It's not really a picture of your mom, but let's say it is.



Cat

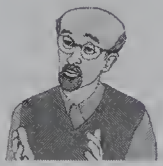
This is a kitty cat. [Show picture of cat] This kitty cat has special eyes that let him see in the dark.

Appendix
(*Continued*)



Heroman

[Show picture of Heroman and say:]
This is Heroman. Heroman has super powers. He can fly very fast so that he can help lots of people all over the world. He *also has eyes* that let him see the insides of things, he can even see through walls.
[Show child a pen] Can you see this pen? [Place pen out-of-sight, behind paper]
Now can you see the pen? [Place picture of Heroman on child's side of the paper]
Well, Heroman can still see the pen. He can see through the paper and see the pen on the other side.
Remember, Heroman can fly very fast and can see right through things.



Mr. Smart

[Show picture of Mr. Smart, facing the child (away from the box) and say:]
This is Mr. Smart. Mr. Smart has special powers. He knows everything.
[Show child closed opaque container that has a ball inside]
Do you know what's inside here? [Child responds: ''No'']
Well, this is the first time that I've played with this, so I don't know what's inside either. Mr. Smart *also* hasn't played with this before. But because he's so smart he still knows what's inside. We would have to look inside, but he wouldn't even need to look.
Mr. Smart, what do you think is in here? [Lean next to Mr. Smart]
Mr. Smart thinks that there is a ball inside. Let's see. [Open container and show child the ball]
Mr. Smart was right! Wow, he knows everything!
Remember, Mr. Smart is very smart. He knows everything.
[Place Mr. Smart face-down, away from box]

God

Let's talk about God. [Show picture] It's not really a picture of God, but let's say it is.

Note. Picture representing God (and language alluding to God's physical presence) was only presented to half of the sample.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1490–1503

Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children

Stacey N. Doan and Qi Wang
Cornell University

This study examined in a cross-cultural context mothers' discussions of mental states and external behaviors in a story-telling task with their 3-year-old children and the relations of such discussions to children's emotion situation knowledge (ESK). The participants were 71 European American and 60 Chinese immigrant mother–child pairs in the United States. Mothers and children read a storybook together at home, and children's ESK was assessed. Results showed that European American mothers made more references to thoughts and emotions during storytelling than did Chinese mothers, who commented more frequently on behaviors. Regardless of culture, mothers' use of mental states language predicted children's ESK, whereas their references to behaviors were negatively related to children's ESK. Finally, mothers' emphasis on mental states over behaviors partially mediated cultural effects on children's ESK.

Children's understanding of emotions has emerged as an important foundation for a range of adaptive behaviors related to social competence and school readiness (Denham, 2006; Izard et al., 2001). The development of the abilities to grasp the causes and consequences of one's own as well as others' emotions begins early. Children start to use basic emotional terms at about 20 months (Bretherton, McNew, & Beeghly-Smith, 1981) and are able to match emotional terms with facial expressions at age 2 (Russell & Bullock, 1986). During the preschool years, children come to be increasingly adept at understanding the situational antecedents of feeling states and can identify individual emotions and the situations that provoke them (Denham, 1986; Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991; Harris, Olthof, Terwogt, & Hardman, 1987). Furthermore, preschoolers have come to understand how perception, desire, and emotion are related in a causal way (Wellman, Phillips, & Rodriguez, 2000), and how thinking about past situations can influence current emotions (Lagattuta & Wellman, 2001).

A burgeoning field of research has focused on early socialization as playing a critical role in children's developing emotional understanding (Denham, 1993; Dunn, Bretherton, & Munn, 1987; Dunn, Brown, & Beardsall, 1991; Eisenberg, Cumberland, & Spinrad, 1998). Specifically, parent–child conversations about emotions and mental states have been shown to be pivotal in predicting children's later understanding of emotion (Dunn, Brown, & Beardsall, 1991; Taumoepeau & Ruffman, 2006, 2008). Most of these studies, however, have focused on children of European descent. Less is known about the socialization of emotional understanding in different cultural contexts and its implications for developmental outcomes. In this study, we examined early narrative practices in European American and Chinese immigrant families in the United States and the impact of these practices on children's developing emotion situation knowledge (ESK).

Socialization of Emotional Understanding

In general, collaborative conversations rich in mentalistic terms with adults, peers, and siblings

---

This research was supported by NIMH Grant R01-MH64661 to Qi Wang. The authors thank members of the Social Cognition Development Lab at Cornell University for their contributions to the project, Francoise Vermeylan for her statistical expertise and advice, and the three anonymous reviewers who provided insightful critiques on earlier drafts of this article. Special thanks go to the children and families who made this study possible.

Correspondence concerning this article should be addressed to Stacey Doan, Department of Human Development, Cornell University, MVR Hall, Ithaca, NY 14853-4401. Electronic mail may be sent to sbd9@cornell.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0013

may aid children's social-cognitive development in various ways. Naming psychological states makes the states explicit, which draws attention to and encourages reflection on these states (Dunn, Brown, Slomkowski, et al., 1991; Ornstein, Haden, & Hedrick, 2004; Wellman, Harris, Banerjee, & Sinclair, 1995). Furthermore, during such conversations children are often exposed to the possibility of differing perspectives and conflicting emotions between the self and others, which may further facilitate their perspective taking and emotional understanding (Harris, 1999).

In line with these theoretical proposals, longitudinal research has shown that early exposures to discussions of mental states contribute to children's emotional understanding (Dunn, Brown, & Beardsall, 1991; Taumoepeau & Ruffman, 2006, 2008). Furthermore, Meins and her colleagues have identified maternal mind-mindedness as a particular type of interaction style that seems to be an important mechanism underlying children's development of complex social cognitive abilities (Meins & Fernyhough, 1999; Meins et al., 2002; Meins et al., 2003). Mothers who are high on mind-mindedness treat their children as autonomous, independent individuals with desires, feelings, and thoughts. During conversations, they pay great attention to their children's mental attributes as opposed to physical appearance or behavioral tendencies. Although mind-mindedness was originally conceptualized as having both behavioral (e.g., looking at the object at which the infant's gaze was directed) and linguistic components (e.g., appropriate references to mental states), further research has shown that the specific aspect of mind-mindedness that appears to facilitate children's social-emotional understanding is the frequency with which mothers refer to mental states, desires, and emotions (Ruffman, Slade, & Crowe, 2002; Taumoepeau & Ruffman, 2006, 2008).

A study by Taumoepeau and Ruffman (2006) provided further evidence for the effects of maternal usage of mental states language on children's emotional understanding. In this study, mothers were asked to describe picture books to their children when children were 15 and 24 months old. Children's emotional understanding was assessed at the second time point by their ability to match emotional faces with a variety of affect-eliciting situations (e.g., a boy being chased by a lion) and by their ability to determine how a person was feeling by their body position. Mothers' references to desires predicted children's later usage of mental state terms, as well as children's emotional understanding. These effects were independent of children's language skills, family socioeconomic status (SES), and mothers' performance on an emotion recognition task. In a follow-up study (Taumoepeau & Ruffman, 2008), the researchers visited the children at 33 months and found that mothers' references to thoughts and knowledge were a more consistent predictor of children's emotional understanding at this later age. The researchers suggest that mothers scaffold children's understanding of mind through their use of specific mental state terms, starting first with desires and progressing to cognitions when children's understanding of mental states are more nuanced.

Thus, family discussions of mental states appear to play an important role in the socialization of emotional understanding. Notably, however, this suite of research has mainly focused on children and families of European descent. At its core, mental states language conveys individual realities and experiences to others, with the emphasis being placed on the internal (e.g., the psychological states of the individual) as opposed to the external (e.g., social and contextual factors). Such an emphasis may be further related to the Western cultural value of individualism, autonomy, and independence. Despite research documenting divergent cultural value systems (Greenfield, Keller, Fuligni, & Maynard, 2003; Kagitcibasi, 2005; Markus & Kitayama, 1991) and their effects on communication styles and preferences (Kim, 2002), there is no study to date that directly examines family discussions of mental states in different cultural contexts and their consequences for children's emotional understanding.

Early Narrative Practices in Cultural Contexts

Narrative practices may reflect the theory of language upheld in a culture (Hansen, 1983, 1989). In Western cultures, language often maps onto ''belief psychology,'' with the function of language being descriptive or representative. In this context, the communicative use of language tends to convey meaning, including ideas, thoughts, and concepts. This is in contrast to traditional Chinese culture where the use of language often serves a regulative function in guiding behavior and coordinating social interactions. As Hansen (1983, p. 61) claims, in the Chinese context ''the function of words is to engender and express attitudes with implications for action rather than to express some 'content' such as the speaker's thoughts.'' Consistent with this view, recent research examining the way in which mothers communicate with their children has found that during personal storytelling, European American parents more frequently encourage children to describe their thoughts, preferences, wishes, and perceptions than Japanese, Korean, and Chinese parents, who are more didactic, focusing on children's behaviors and moral obligations (Martini, 1998; Wang, 2001; Wang, Leichtman, & Davies, 2000).

Particularly relevant to the current research, Wang et al. (2000) found that compared with Chinese mothers, European American mothers often made more comments on mental states and were more likely to refer to the child's personal needs, preferences, judgments, and opinions when sharing memories and telling stories. These cultural differences in mother–child conversations appear to play out even when the subject under discussion is emotional. For example, Fivush and Wang (2005) examined Chinese and European American mothers' conversations about positive and negative emotionally salient events (e.g., birthday parties, conflict with parents or peers). They found that although the sheer volume of conversations did not differ between cultures, American mothers provided more causal explanations for both their children's and others' emotions than did Chinese mothers, who tended to emphasize social norms and behavioral expectations when discussing emotions with their children.

These different beliefs and practices pertaining to the functions of language in communicative contexts are further echoed in the value systems across cultures. In European American culture where a great emphasis is placed on individuality and agency, individuals are encouraged to express their thoughts, ideas, and emotions, which, in turn, help to affirm their autonomous sense of self. In contrast, Chinese culture prizes group harmony and interrelatedness; emphasis is placed on an individuals' overt behavior and its regulation because it may have direct consequences on group functioning. Individuals' thoughts and emotions, in this context, are often deemphasized as they may be socially divisive (Markus & Kitayama, 1991). Indeed, research has shown that East Asian adults and children rely more on behavior-descriptive verbs in person descriptions than their Western counterparts, who tend to prefer trait adjectives (Maas, Kurasawa, Politi, & Suga, 2006; Wang, 2004). East Asian children have been shown to justify another person's false beliefs by overt behaviors and social rules rather than mental states (Naito & Koyama, 2006). Furthermore, Chinese parents tend to view their children's behavior as highly malleable rather than genetically determined, which leads to their emphasis on focusing and regulating behaviors in the child (Stevenson, Chen, & Lee, 1993).

The cultural difference in emphasizing mental states versus external behaviors may have important consequences for children's developing emotional understanding. When asked to judge the emotional nature of story situations or to describe situations likely to provoke various emotions, European American preschoolers scored higher than their Chinese peers regardless of age, and they made more rapid progress in ESK over time (Wang, 2003; Wang, Hutt, Kulkofsky, McDermott, & Wei, 2006). Conceivably, the emphasis on mental states and affect in early narrative practices in European American families may have facilitated European American children's emotional understanding. Nonetheless, there has been no direct evidence concerning the influence of maternal references to mental states on children's developing emotional understanding across cultures. Also, no cross-cultural data are available about family discussions of external behaviors and how they may be related to children's emotional understanding. Conceivably, a focus on external behaviors may draw children's attention outward and render inner psychological states less salient. As a result, it may hinder the development of emotional understanding.

Purposes of the Current Study

The purpose of this study is threefold. First, we examine references to mental states and overt behaviors during mother–child storytelling in European American and Chinese immigrant families. Second, we examine the relation of mental states and behavioral references to children's emotional understanding. We focus on children's understanding of situations that elicit specific emotions, namely, ESK (Denham, Zoller, Couchoud, & Holt, 1990; Harris et al., 1987). This aspect of emotional understanding may be critical for the actual emotional experience within event situations (Fridja, 2008) and may play an important role in affecting how children understand and represent significant personal experiences (Wang, 2008). Furthermore, the ability to understand how situational cues can be used to infer others' emotions is crucial for children's navigation of social relationships and their social-emotional well-being (Fine et al., 2003; Izard et al., 2001).

Finally, we test a mediation model examining whether the emphasis on mental states rather than behaviors during mother–child storytelling mediates cultural influences on children's ESK. We hypothesized that during storytelling, Chinese mothers would make more references to behaviors than European American mothers, who would make more references to mental states. And consistent with previous findings (Wang, 2003; Wang et al., 2006), European American children would score higher on ESK than Chinese children. Furthermore, maternal emphasis on mental states relative to behaviors in conversations would have a facilitative effect on children's ESK, regardless of culture. Finally, we hypothesized that mental states language would serve as an important mediator that gave rise to cultural differences in children's ESK.

Method

Participants

The participants were 60 Chinese immigrant (30 boys; age range = 30–43 months, $M = 35.00$ months) and 71 European American (37 boys; age range = 30–42 months, $M = 35.49$ months) children and their mothers from a university town and suburban areas in upstate New York. The mean age was not significantly different across the two cultures. Children were recruited through local schools and by word of mouth, and were taking part in a larger longitudinal study of sociocognitive development across the preschool years. All children came from middle-class backgrounds, with the majority of the mothers (Chinese immigrant, 98%; European American, 93%) having obtained a college degree or beyond. Chinese immigrant families were originally from mainland China, Hong Kong, and Taiwan, with the majority from mainland China (93%). Most (80%) of the children were born in the United States.

Procedure

Two female researchers visited mothers and children in their homes. Chinese-English bilingual researchers visited the Chinese immigrant families. Of the 60 Chinese immigrant children, 80% spoke only Chinese and the rest spoke English or a mixture of English and Chinese. Before commencing with children's interviews, the researchers established familiarity and rapport with the children by chatting about nonrelevant events. All materials were written in both English and Chinese and a translation and back-translation procedure was carried out to ensure their equivalence in both literal and sense meaning. Mothers were asked to engage the child in a series of free play and semistructured tasks, followed by a researcher–child session. The entire home visit took approximately 1½ hr and was video tape-recorded. Only the tasks relevant to the current study are described here.

Measures

Mother–child storytelling. Mothers' use of mental states language and references to behaviors were assessed in a storytelling task. Joint picture book reading has been found to be a common practice among middle-class families in both European American and Chinese cultures (Johnston & Wong, 2002; Li, 2006). Mothers received a book entitled *Bear Goes to the Market*, a children's book with illustrations but no words. The storybook is 17 pages long and contains culturally neutral themes. It has been developed for and has been used in previous cross-cultural research with European American and Chinese samples (Han, Leichtman, & Wang, 1998; Wang et al., 2000). The story depicted a mother bear and a baby bear dropping off mail, going to the grocery store (where the baby bear briefly loses his mother and gets upset because he wants a cake but is not allowed to have it), and playing in a playground. Mothers were asked to tell the story with their children and to make up the text as they went along. They were also told that they could tell the story in any way they would like and take as long as they wanted to. For Chinese mother–child dyads, time spent on the storytelling task ranged from 3.35 to 15.35 min ($M = 7.29$ min, $SD = 2.79$). For European American mother–child dyads, it ranged from 2.03 to 18.52 min ($M = 7.09$ min, $SD = 3.25$). There was no difference in the mean length of the storytelling between the two cultural groups.

Coding of the story narratives was done in the original language. A native English speaker coded the American data and a native Chinese speaker coded the Chinese data. Repeated training sessions were held to ensure that both coders were applying the same definitions of the coding scheme to the two data sets. A third coder, who was bilingual and fluent in both Chinese and English, independently coded 20% of each data set for reliability estimate. Any disagreements were resolved through discussion.

Following Ruffman et al. (2002), mothers' references to cognitions (e.g., think, know), desires (e.g., want, like, love), modulations of assertions (e.g., might, maybe), emotional states (e.g., sad, pleased), and other mental states (e.g., remember, decide) were separately tabulated. These references were then combined to form a composite category of *mental states language*. Coding for the Chinese data was further informed by research examining mental states language in Chinese children (Lu, Su, & Wang, 2008; Tardif & Wellman, 2000). Terms were counted as referring to mental states when the usage was judged, in context, to be indicative of desires (e.g., 要 yao4), emotions (e.g., 生气 sheng1qi4), cognitions (e.g., 想 xiang3, 知道 zhi1-dao4), other mental states (e.g., 记得 ji4de), and modulations of assertions (e.g., 可能 ke2neng2). Terms that could be interpreted as behavior or mental states (e.g., 看 kan4, equivalent to "see" in English, but may be understood as "think" in Chinese) were coded based on the context. Intercoder reliability indexed by Cohen's kappas ranged from .96 to 1.00 for the Chinese data ($M = 0.98$) and from .87 to .96 for the American data ($M = 0.93$).

Mothers' references to overt behaviors, including physical actions depicted in the book (e.g., running, playing) and potential actions that were not necessarily seen in the book (e.g., he's going to go home now) were coded as *behavioral descriptions*. Cohen's kappas were .83 for the Chinese data and .95 for the American data. This variable further included a subcategory of mothers' references to the behavioral and descriptive manifestations of emotions (e.g., "The bear is crying" and "aw, the bear has tears on his face"). Compared with references to internal emotional states (e.g., "The bear is sad"), behavioral descriptions of emotions focused on external or physical aspects of emotions (Ruffman et al., 2002). Cohen's kappas for this subcategory were .98 for the Chinese data and .87 for the American data. Finally, a "mental-state-to-behavior" ratio was calculated for each mother to form a composite score of *mental states focus*, which controlled for possible influence of narrative length. Children's references to mental states and external behaviors were also coded. However, the frequencies were too low to warrant analysis (see also Slaughter, Peterson, & Mackintosh, 2007) and therefore not considered further.

Language ability. Mothers filled out a shortened version of the Child Development Inventory (Ireton, 1992) that assessed children's language skills. Both expressive language and language comprehension were assessed (possible score range = 0–100; Cronbach's $\alpha = .93$). The questionnaire was translated and back-translated by two English–Chinese bilingual research assistants and checked by a native Chinese speaker to ensure a balance of literal and sense meaning, as well as natural sounding expression. English words and grammatical rules that were not used in Chinese were exchanged for the Chinese equivalents. The inventory was then pilot-tested with seven European American families and five Chinese immigrant families to ensure equivalence. Because children's expressive language score and language comprehension score were highly correlated, $r = .80$, $p < .0001$, they were combined for later analysis. The language measure was included because language ability has been shown to be related to the understanding of emotions and mental states (Happé, 1995).

Maternal education level. As a measure of SES, maternal education was coded on a 3-point scale: 1 = *high school qualification*, 2 = *4-year college degree*, and 3 = *postgraduate degree*.

Children's ESK. Children's ESK was assessed using an emotion production task adapted from previous research examining in different cultures children's ability to understand situations that provoke various emotions (Harris et al., 1987). This task has been used with European American and Chinese preschoolers (Wang, 2008; Wang et al., 2006) and has been shown to be effective in eliciting children's responses regarding the situational antecedents of emotions. Children were asked to describe situations likely to provoke fearful, sad, angry, and happy emotions. The researcher presented each term to the child and asked the child to describe situations that would elicit such an emotion in people (e.g., "What makes people feel sad?") and in the child (e.g., "What makes you feel sad?"). For each question, the researcher prompted the child to provide as many situations as possible (e.g., "What else makes people feel sad?") until the child indicated that that he or she was finished. The researcher rewarded children with stickers to motivate them to come up with as many situations as possible. This task took approximately 10 min.

The number of responses children provided was tallied. Responses were considered correct if the situation described by the child was judged to be able to elicit the presented emotion (e.g., "What makes you feel happy?" "My teddy bear"). This included situations that would elicit the emotion in a child but not necessarily in an adult. In both cultural groups, responses provided by children were mostly correct. Incorrect responses were rare, whereby there was on average less than one incorrect response in each group. On the other hand, not all children could provide a correct response to every emotion. In a few cases, children randomly named objects in the room (e.g., "What else makes people feel happy?'' ''table, chair, cup, . . .''), and these responses were not coded.

The number of correct responses children provided for the four emotions was summed to form a composite score of ESK. A second research assistant coded 20% of the data for reliability. Cohen's kappa, calculated for each child, ranged from .71 to 1.00 ($M = 0.84$) for the Chinese children and from .66 to 1.00 ($M = 0.75$) for the European American children.

Results

Preliminary Analyses

Analyses of mother–child storytelling variables were based on frequencies. Previous research on mother–child interaction has shown that compared with proportion, frequency is more informative in reflecting the sheer amount of different types of information that children are exposed to and is also more predictive of child outcome (e.g., Fivush, 1998; Han et al., 1998; Reese et al., 1993). Ruffman et al. (2002) further argue that each utterance potentially has a direct bearing on children's understanding of mental states because it provides additional input, just in the same way that having multiple older siblings or spending more time with older individuals has a beneficial effect on children's theory of mind (Lewis et al., 1996; Ruffman et al., 1998). In addition, in studies where both frequency and proportion were analyzed, the analyses yielded similar patterns of results (Meins, 2003; Meins et al., 2003; Wang, 2006).

Nevertheless, to ensure that the number of maternal references did not simply reflect overall maternal involvement or verbosity, and hence contributing to children's ESK, we examined the sum of behavioral and mental states references ($M = 63.28$, $SD = 27.08$) as well as the total amount of time spent on the storytelling task ($M = 7.18$, $SD = 3.04$) in relation to children's ESK. Results revealed no significant correlation between the sum of references ($r = .10$, $p = .26$) or the total time spent on the task ($r = .09$, $p = .29$) and children's ESK, thus ruling out the possibility that the overall linguistic input was influencing children's ESK. In addition, there was no significant cultural difference in mothers' total references to mental states and behaviors combined (European American: $M = 66.43$, $SD = 29.01$; Chinese: $M = 59.49$, $SD = 25.26$), $F(1, 128) = 2.14$, $p = .15$, nor in the amount of time mother–child pairs spent on the storytelling task (European American: $M = 7.02$, $SD = 3.15$; Chinese: $M = 7.30$, $SD = 2.79$), $F = (1, 128) = .38$, $p = .50$. Nonetheless, to ensure that cross-cultural differences were not driven by overall maternal involvement or verbosity, we included the amount of time spent on the task as a covariate in analyses wherever appropriate.

Scores for both the composite mental state and behavioral references were positively skewed. Logarithmic transformations were thus performed on these variables before they were submitted to analysis. To facilitate comparisons with previous research, untransformed means are reported.

The following results are presented in accordance with our three hypotheses. The first group of analyses examined group (culture and gender) and individual variations (child age, language skills) in maternal references to mental states and behaviors, as well as in children's ESK. The second set of analyses aimed at determining the relation between types of maternal references (i.e., mental states and behaviors) and children's ESK, independent of other group and individual factors. The final set of analyses tested the mediational role of a mental states focus for cultural influences on children's ESK. Some of the children did not complete all the tasks; thus, the degrees of freedom varied slightly across tests.

Maternal References to Mental States and Behaviors and Child ESK

Table 1 lists the means and standard deviations of mother and child variables by culture. To test our first hypothesis that European American mothers would focus more on mental states whereas immigrant Chinese mothers would focus more on behaviors, we first ran a 2 (culture) × 2 (gender) multivariate analysis of covariance (MANCOVA) on the subtypes of maternal references to mental states (e.g., desires, cognitions), with child age, language ability, maternal education, and time spent on the task as covariates. Consistent with our prediction, the effect of culture was highly significant, $F(6, 116) = 15.67$, $p < .001$, $\eta_p^2 = .45$. There was also an effect of time, $F(6, 116) = 14.19$, $p < .001$, $\eta_p^2 = .42$, suggesting that more time spent on the task was associated with increased maternal input. No other main effects or interactions reached significance.

We then ran separate ANCOVAs to determine the effect of culture on each type of mental state input. As expected, European American mothers made significantly more references to cognitions

Table 1
Means and Standard Deviations of Mother and Child Variables by Culture

	Chinese immigrant		European American	
	M	*SD*	*M*	*SD*
Child variables				
Age (months)	35.00	3.43	35.49	3.30
Language***	71.15	16.44	82.48	11.64
Emotion situation knowledge***	3.77	5.19	13.86	8.58
Mother variables				
Education	2.65	0.52	2.52	0.63
Mental states language***	10.25	7.12	29.76	18.11
Cognitions***	3.44	4.47	17.46	14.81
Desires	3.40	3.41	4.23	4.00
Modulations of assertions***	0.95	1.92	3.58	3.31
Emotional states**	1.74	2.21	3.11	2.71
Other mental states***	0.73	1.60	1.70	1.47
Behavior descriptions***	49.23	20.45	36.68	14.60
Emotional behavior***	2.73	2.27	1.00	1.15
Mental states focus***	0.21	0.12	0.84	0.59

**$p < .01$. ***$p < .001$ (all significance tests are one-tailed).

(think/know), $F(1, 122) = 54.20$, $p < .001$, $\eta_p^2 = .31$; modulations of assertions, $F(1, 122) = 29.97$, $p < .001$, $\eta_p^2 = .20$; emotional states, $F(1, 122) = 5.28$, $p = .02$, $\eta_p^2 = .04$; and other mental states, $F(1, 122) = 10.10$, $p = .002$, $\eta_p^2 = .08$, than did Chinese immigrant mothers. There was no cultural difference in the number of references to desires, $F(1, 122) = 1.23$, $p = .27$, $\eta_p^2 = .01$. There was a main effect of time for references to cognitions, $F(1, 122) = 31.57$, $p < .001$, $\eta_p^2 = .20$; desire, $F(1, 122) = 21.83$, $p < .001$, $\eta_p^2 = .15$; modulations of assertions, $F(1, 122) = 23.14$, $p < .001$, $\eta_p^2 = .16$; emotional states, $F(1, 122) = 6.63$, $p = .01$, $\eta_p^2 = .05$; and other mental states, $F(1, 122) = 6.00$, $p = .02$, $\eta_p^2 = .05$. There were no significant effects or interactions pertaining to gender, age, language, and maternal education.

Consistent with our hypotheses, analysis of the composite mental states language score showed that even after controlling for mothers' education level, time spent on task, child language ability, child age, and gender, European American mothers made significantly more references to mental states than did Chinese mothers, $F(1, 122) = 78.02$, $p < .001$, $\eta_p^2 = .39$.

Next, we analyzed the overall score of behavioral descriptions, with the same covariates. Results showed that Chinese immigrant mothers made more references to behaviors than did European American mothers, $F(1, 122) = 24.83$, $p < .001$, $\eta_p^2 = .16$. Chinese immigrant mothers were also more likely to describe more emotional behaviors than did European American mothers, $F(1, 122) = 23.43$, $p < .001$, $\eta_p^2 = .16$. European American mothers scored higher on mental states focus than did Chinese immigrant mothers, $F(1, 122) = 56.76$, $p < .001$, $\eta_p^2 = .32$. Finally, it is important to note that across both cultures, references to behaviors were more common than references to mental states.

With regard to children's ESK, a 2 (culture) × 2 (gender) ANCOVA with child age and language ability as covariates showed that European American children scored significantly higher than did their Chinese peers, $F(1, 127) = 48.15$, $p < .001$, $\eta_p^2 = .28$. No effects pertaining to gender, age, and language neared significance.

Relation of Maternal References to Mental States and Behaviors to Children's ESK

The second goal of the current study is to examine the relation of maternal mental states language and behavioral references to children's ESK. Zero-order correlations were first calculated to examine the interrelations among maternal references to mental states and behaviors and children's ESK score for each culture. Table 2 reports the zero-order correlations among mother and child variables for each culture. For the European American sample, ESK was positively related to mothers' references to cognitions, the sum total of mental states, and mental states focus. For the Chinese sample, ESK was positively related to total mental states references and mental states focus; albeit the correlations did not reach significance. In addition, ESK was negatively correlated with maternal references to behaviors in the European American sample and negatively correlated with references to emotional behaviors in the Chinese sample. Language was not significantly correlated with ESK in either sample.

Partial correlations were further calculated between maternal variables and children's ESK, controlling for culture, child age, gender, language, maternal education, and time spent on task. Table 3 reports the partial correlations. Independent of the group and individual variables, maternal references to cognitions and maternal mental states focus were positively correlated with children's ESK. In addition, maternal references to behaviors were negatively related to children's ESK score.

Next, *hierarchical* regression analyses were conducted to examine the effects of mental states

Table 2
Zero-Order Correlations Among Mother and Child Variables by Culture

	1	2	3	4	5	6	7	8	9	10	11	12	13
European Americans													
1. Age	–	.37**	–.05	.08	.09	.13	.05	–.13	–.01	.12	.05	–.08	.14
2. Language skills	.31**		.03	.20*	.03	.10	.05	.10	.18	.07	.12	–.02	.07
3. ESK	–.02	.16		.05	.22*	.02	.09	–.07	.13	.19*	–.23*	.01	.24*
4. Mother's education	.03	–.06	–.08		.06	.02	–.02	–.20*	–.08	.03	–.15	.15	.15
5. Cognitions	.11	.06	–.09	–.06		.26*	.41***	.21*	–.16	.94***	–.07	.02	.72***
6. Desires	.38**	.14	–.06	–.25*	.00		.10	.24*	.27*	.47***	.24*	–.08	.12
7. Modulations of assertions	.08	–.06	.10	–.01	.03	.01		.02	.15	.54***	.18	.08	.26*
8. Emotional states	–.01	.17	.12	.03	.04	.07	.02		.17	.34**	.08	–.19	.06
9. Other mental states	.08	–.04	–.06	–.16	.17	.07	–.09	.37**		.27*	.15	.04	.13
10. All mental states	.29*	.13	.10	–.19	.69***	.52***	.27*	.46***	.45***		.06	.01	.68***
11. Behavior descriptions	–.04	.23*	–.06	–.15	.02	.33**	.05	.16	.11	.26*		–.03	–.37**
12. Emotional behavior	–.14	–.04	–.24*	–.05	.11	.19	–.07	.23*	.12	.24*	.28*		.03
13. Mental states focus	.34**	.09	.11	–.07	.64***	.29*	.38**	.32**	.37**	.81**	–.20	–.09	
Immigrant Chinese													

Note. European American correlations are above the diagonal, and immigrant Chinese are below. ESK = emotion situation knowledge.
*$p < .05$. **$p < .01$. ***$p < .001$ (all significance tests are one-tailed).

Table 3
Partial Correlations Between Mother References to Mental States and Behaviors and Children's Emotion Situation Knowledge (ESK)

Mothers' utterances	Child's ESK
Cognitions	.15*
Desires	–.04
Modulations of assertions	.06
Emotional states	–.05
Other mental states	.03
Behavior descriptions	–.18*
Emotional behavior	–.14
Mental states focus	.20*

Note. Partialing out culture, mother's education, child's language abilities, age, time, and gender.
*$p < .05$ (all significance tests are one-tailed).

Table 4
Hierarchical Regression Analyses for Variables Predicting Children's Emotion Situation Knowledge (ESK)

		ESK			
Model	Variable	*B*	*SE*	β	R^2 and ΔR^2
1	Culture***	9.81	1.41	0.56	.34***
	Gender	–0.90	1.33	–0.05	
	Age	–0.16	0.20	–0.06	
	Language skills	0.04	0.05	0.07	
	Mother's education	0.31	1.12	0.02	
	Time spent on task	0.23	0.22	0.08	
2	Culture**	5.20	2.05	0.30	.05**
	Gender	–1.38	1.29	–0.08	
	Age	–0.31	0.20	–0.18	
	Language skills	0.05	0.05	0.08	
	Mother's education	–0.08	1.09	–0.01	
	Time spent on task	0.26	0.27	0.09	
	Mental states language*	2.67	1.12	0.28	
	Behavioral descriptions**	–5.70	1.92	–0.29	

Note. Immigrant Chinese and females were set as the reference group.
*$p < .05$. **$p < .01$. ***$p < .001$.

language (composite score) and behavioral references on children's ESK, independent of culture, gender, age, language skills, time spent on task, and maternal education. Table 4 reports the summary of results from the regression analyses. We were interested in whether maternal inputs could account for variance in children's ESK independent of group and individual differences. Therefore in the first step of the regression, demographic variables including culture, gender, age, language, and maternal education, along with time spent on task were first entered. These variables combined predicted 34% of the variance in children's ESK, $F(6, 123) = 10.65$, $p < .001$. Next, to see if maternal input would increase the variance explained above these control variables, we added maternal references to mental states and behaviors to Step 2 of the model. Maternal references to mental states and behaviors increased the variance explained to 40%, $F(8, 121) = 9.87$, $p < .001$. Finally, to test whether culture moderated the effects of mental states language and behavioral references, we added the interaction terms of Culture × Mental States and Culture × Behaviors to the model. Neither interaction reached significance, suggesting that the effects of internal states and behaviors were not moderated by culture.

A second set of regression analyses was conducted to examine whether the proclivity to focus on mental states over behaviors (i.e., mental states focus) would predict children's ESK. In the first step of the regression, demographic variables including culture, gender, age, language, and maternal education, along with time spent on task, were first entered. Mental states focus was then added in the second step. Mental states focus significantly predicted children's ESK, $\beta = .22$, $F(6, 123) = 10.22$, $p < .001$, and significantly increased the proportion of variance explained in ESK, $\Delta R^2 = .03$, $F(1, 122) = 10.22$, $p = .02$. Again, we found no moderation effect of culture.

In sum, the results indicated that regardless of culture, maternal use of mental states language positively predicted children's ESK whereas maternal references to overt behaviors were negatively related to children's ESK.

Did Mental States Focus Mediate Cultural Effects on ESK?

Our final set of analyses tested the hypothesis that mothers' proclivity to focus on mental states rather than behaviors would mediate the cultural effect on children's ESK. All regression models controlled for child gender, age, language, maternal education, and time spent on task. The mediation results are summarized in Figure 1. First, to establish conditions for mediation (Baron & Kenny, 1986), culture (i.e., the independent variable) was examined as a predictor of children's ESK (i.e., the criterion); the result was significant, $F(6, 122) = 10.65$, $p < .001$. European American mothers also scored higher than Chinese mothers on mental states focus (i.e., the mediator), $F(6, 123) = 12.62$, $p < .001$, and mental states focus positively predicted children's ESK independent of culture, $F(6, 123) = 7.3$, $p < .001$. Finally, when both mental states focus and culture were entered into the model as predictors, the effect of culture on ESK was reduced in size, although still significant, $p < .001$. The reduction was significant by the stringent Sobel's (1982) test for mediation, $Z = -4.29$, $p < .001$. In sum, maternal mental states focus mediated, at least partially, the effect of culture on children's ESK.

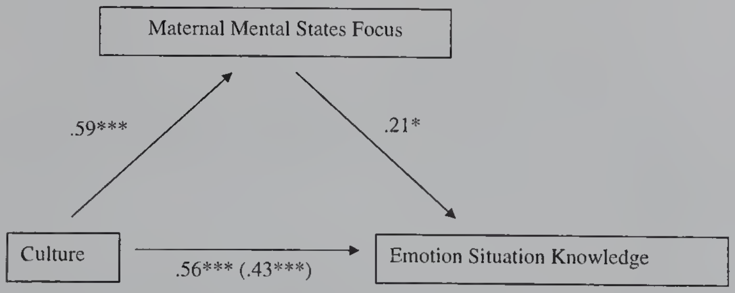


Figure 1. Mental states focus mediated the effect of culture on children's emotion situation knowledge.
Note. Regression models controlled for child gender, age, language, mothers' education, and time spent on task. The coefficient in parentheses was from the regression when the mediator (mental states focus) was included in the equation.
$^{*}p < .05$. $^{***}p < .001$.

Discussion

The present study is the first to examine, in a cross-cultural context, maternal references to mental states and external behaviors during narrative interactions with their children and the relations of these maternal references to children's ESK. The results revealed varied narrative practices across cultures, which further differentially affected ESK in children. As predicted, European American mothers used significantly more mental states language than did Chinese immigrant mothers, whereas Chinese immigrant mothers made more references to behaviors than did European American mothers. The differences held true even when mothers were talking about emotions; European American mothers referred more frequently to internal emotional states, whereas Chinese immigrant mothers described more physical and behavioral manifestations of emotions.

The results are consistent with previous research demonstrating that during conversations with their young children, Chinese mothers often place an emphasis on external behaviors and actions, whereas European American mothers tend to focus on the child's internal attributes, thoughts, and feelings (Fivush & Wang, 2005; Wang, 2001). European American mothers' tendency to emphasize thoughts and feelings may reflect the cultural assumption that language is a mechanism for expression, specifically of one's mental states, whereas Chinese mothers' focus on behaviors may exemplify the importance of language as a mechanism for "guiding actions and social interaction" (Hansen, 1983, p. 85). In addition, the emphasis on mental states places importance on individual realities and states of being, reflecting European American cultural values of individuality and autonomy (Wang et al., 2000). On the other hand, behavior-descriptive terms emphasize individuals' explicit actions and relations to the physical and social contexts (Maas et al., 2006).

Notably, although there were cultural differences in most aspects of mental states references, there was no difference in maternal references to desires. Although desires are considered a mental state, they have clear implications for behaviors. Because of this advantage, children acquire verbs for desires before they acquire verbs for thoughts and beliefs (Bartsch & Wellman, 1995), and by the early preschool age children have demonstrated an understanding of simple desires (Repacholi & Gopnik, 1997; Wellman et al., 2000). Given that desire terms have important implications from both a psychological and a behavioral perspective, it makes sense that mothers from both cultures would readily use desire terms in conversations with their young children. Interestingly, research examining Chinese children's use of mental states language has shown that although they acquire desire terms earlier than their European American peers, they lag behind in terms related to cognitions (Tardif & Wellman, 2000). Furthermore, there is some evidence to suggest that Chinese children's early vocabulary contains more verbs than nouns (Choi & Gopnik, 1995; Tardif, 1996). These findings may be explained by the Chinese pattern of maternal input found in our study. The maternal emphasis on behaviors and desires may facilitate Chinese children's learning of verbs and desire terms, respectively, whereas the lack of maternal references to cognitions may impede the children's acquisition and hence expression of thoughts and beliefs.

As expected, across the cultural groups, mothers' use of mental states language was positively associated with children's ESK, consistent with previous research with Western samples (Taumoepeau & Ruffman, 2006, 2008). Furthermore, our results showed that maternal references to behaviors were negatively associated with children's ESK. Indeed, mothers' proclivity to focus on mental states as opposed to behaviors partially mediated the cultural effect on children's ESK. Although mental states language encourages children to reflect on thoughts, desires, and feelings of themselves and others and thus facilitates their understanding of cognitive and affective states (Symons, 2004), behavioral descriptions may draw children's attention to external and contextual factors and therefore may hinder their emotional understanding. Furthermore, the emphasis on mental states argues for the notion that individual behavior can be understood in terms of inner desires, emotions, and thoughts, and that people can be understood in terms of their psychological states (Flavell, 1988). By using language to make explicit the connections between the internal and the external, maternal references to mental states may aid in children's understanding of the psychological causes and antecedents of emotions. References to behaviors are more likely to draw attention to physical realities and social relations and thus may undermine the importance of psychological states, including one's own and others' emotions.

The differential focus on mental states versus behaviors in socialization practices may reflect cultural variations in folk psychology (Lillard, 1998; Miller, 1984). In Western cultures, the locus of the self is thought to lie in mental attributes of the person, such as attitudes, preferences and motives, and mental states are considered important means of understanding others' behaviors (Choi, Nisbett, & Norenzayan, 1999). In contrast, Eastern cultures place the self in context and stress the importance of proper behaviors and actions based on interpersonal norms and social rules (Shweder, 1998). The different narrative practices found in our study suggest that children's attention is directed toward different aspects of the individual, with Chinese mothers emphasizing behaviors whereas European American mothers stressing psychological states. It is important to note, however, that cultural differences in the emphasis on internal versus external attributes are relative rather than absolute—mothers in both cultures used more behavioral descriptions than mentalistic descriptions.

Consistent with a host of recent studies (Adrian, Clemente, & Villanueva, 2007; Ensor & Hughes, 2008; Taumoepeau & Ruffman, 2008), we found maternal references to cognitions to be the strongest correlate with preschool children's emotional understanding. Also similar to some previous work (Ensor & Hughes, 2008; Taumoepeau & Ruffman, 2006, 2008), our study did not find a significant link between maternal references to emotions and children's emotional understanding. We speculate that this might be related to the particular task context involved in the studies. In particular, our book-reading task was not inherently emotional in nature, and mothers might focus on children's thoughts and opinions concerning the story happening rather than their emotions. Indeed, mothers in our study referred to cognitions more than four times as often as emotions. Thus, references to cognitions may be more indicative of maternal mind-mindedness in this particular context (Meins et al., 2002) and is therefore more related to children's emotional understanding. Additionally, in the current study maternal mental state language in reference to emotions mainly captured mothers' labeling of emotions. It is possible that explicit discussion of the causes of feeling states is more effective in helping children understand the situational antecedents of emotions. Future research should directly examine the effect of different contexts in which specific mental state references are used, as well as the different *ways* that parents use mental state language. This will help to further our understanding of the contribution of mental state language to children's emotional understanding.

Results from the current study further demonstrate that the relative emphasis on mental states versus behaviors in mother–child narrative interactions was an important mechanism underlying cultural differences in children's ESK. Nevertheless, the mediation was only partial, and culture remained a highly significant contributor to children's ESK after maternal mental states focus was controlled. Thus, although maternal mental states language may aid children's emotional understanding, it may be only one of a host of factors that may play important roles. For example, research has demonstrated cross-cultural differences in emotional experience including intensity, expression, and evaluation of emotions (Matsumoto, Kudoh, Scherer, & Wallbott, 1988; Scherer, Matsumoto, Wallbott, & Kudoh, 1988). There have been very few studies examining how these differences may play out in socialization practices and further affect children's developing emotional understanding.

Furthermore, our study only utilized one ESK task and that required verbal responses. Perhaps Chinese children were unable or unwilling to speak out their thoughts and, as a result, cultural differences in children's performance on the task might reflect their expressiveness or willingness to communicate rather than their actual emotional understanding. This interpretation is unlikely, however, given that past research has shown that Chinese children are as verbal as European American children in tasks that require self-expression (Han et al., 1998; Wang, 2004; Wang & Leichtman, 2000). In addition, controlling for language ability in analysis did not change the pattern of results. Nevertheless, it is important for future cross-cultural studies to use a variety of emotion tasks, including the ones that require only behavioral responses. Also notably, knowledge of emotion is a broad term involving a host of factors, and we examined only one aspect. In a recent review, Pons, Harris, and de Rosnay (2004) point out that there are at least nine distinct components of emotion knowledge in relation to emotion recognition, perception, and regulation. It is unclear from extant data the extent to which mothers' mental states language has an effect on these various components. Furthermore, the type of maternal input that may be useful is likely to vary as a function of developmental period. For example, maternal emotional expression may be more important than linguistic inputs (such as mental states language) for early understanding of emotions during infancy because of limited language abilities in infants (Friend, 2001). Future research should examine possible interactions among types of maternal input, periods of development, and how they relate to different aspects of emotional understanding in children. In addition, research in the area of mental states language would benefit by examining the extent to which maternal references to mental states may have an effect on children's social cognition (e.g., theory of mind and emotional understanding) or a broad effect on cognitive abilities in general.

It should also be noted that our study only examined one time point. The data were therefore correlational in nature, and causal conclusions cannot be drawn with regard to the influences of maternal use of mental states language and behavioral references on children's ESK at either individual or cultural level. In addition, previous research suggests that mothers tend to increase their references to cognitions as children get older, thus pitching their talk to the developmental level of the child (Taumoepeau & Ruffman, 2008). Although it is possible, then, that mothers (or Chinese mothers as a group) made fewer references to cognitions because their children (or Chinese children as a group) had lower level of general cognitive abilities or language skills, this interpretation is unlikely given that maternal references to cognitions were not related to child age or language skills in the current sample and that they were positively related to ESK independent of child age and language. Additionally, previous research examining maternal talk with preverbal infants (Meins et al., 2002) and studies that controlled for children's conversational input (Ruffman et al., 2002) have shown that the type of maternal input cannot be entirely attributed to child characteristics. Nevertheless, the bidirectional influence between mother and child cannot be completely ruled out from the current data. Further research with longitudinal approaches and experimental designs is called for. Such research will not only help to establish causality but also reveal trajectories of different mental states language usages as children develop in different cultural contexts.

Finally, it is important to further examine the effects of differential socialization emphases on immigrant children straddling two cultures. How does acculturation influence the ways in which immigrant parents interact with their children? What are the sociocognitive consequences of being raised in a family context that is different from the mainstream? These questions will help us further understand the dynamic processes of human development as taking place within multiple contexts and settings where children actively acquire cultural knowledge and eventually become competent members of their society.

References

Adrian, J. E., Clemente, A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*, 1052–1067.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition* (pp. 333–373). Hillsdale, NJ: Erlbaum.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*, 47–63.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language, 22*, 497–529.

Denham, S. A. (1986). Social cognition, prosocial behavior and emotion in preschoolers: Contextual validation. *Child Development, 57*, 194–201.

Denham, S. A. (1993). Socialization of emotion: Pathways to preschoolers' emotional and social competence. *Journal of Nonverbal Behavior, 17*, 205–227.

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, 17*(1), 57–89.

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of peer status in young preschoolers. *Child Development, 61*, 1145–1152.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology, 23*, 132–139.

Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others emotions. *Developmental Psychology, 27*, 448–455.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*, 1352–1366.

Eisenberg, N., Cumberland, A., & Spinrad, T. (1998). Parental socialization of emotion. *Psychological Inquiry, 9*, 241–273.

Ensor, R., & Hughes, C. (2008). Content or connectedness? Mother-child talk and early social understanding. *Child Development, 79*, 201–216.

Fine, S. E., Izard, C. E., & Trentacosta, C. J. (2003). Emotion situation knowledge in elementary school: Models of longitudinal growth and preschool correlates. *Social Development, 15*, 730–751.

Fivush, R. (1998). The stories we tell: How language shapes autobiography. *Applied Cognitive Psychology, 12*, 483–487.

Fivush, R., & Wang, Q. (2005). Emotion talk in mother–child conversations of the shared past: The effects of culture, gender, and event valence. *Journal of Cognition and Development, 6*, 489–506.

Flavell, J. (1988). Developing theories of mind. In D. Olson, J. Astington, & P. L. Harris (Eds.), *Developing theories of mind* (pp. 244–267). Cambridge, UK: Cambridge University Press.

Friend, M. (2001). The transition from affective to linguistic meaning. *First Language, 21*, 219–243.

Frijda, N. H. (2007). *The laws of emotion*. Mahwah, NJ: Erlbaum.

Greenfield, P. M., Keller, H., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 34*, 701–713.

Han, J. J., Leichtman, M. D., & Wang, Q. (1998). Autobiographical memory in Korean, Chinese and American children. *Developmental Psychology, 34*, 701–713.

Hansen, C. (1983). *Language and logic in ancient China*. Ann Arbor: University of Michigan Press.

Hansen, C. (1989). Language in the heart-mind. In R. E. Allinson (Ed.), *Understanding the Chinese mind: The philosophical roots* (pp. 77–124). Oxford, UK: Oxford University Press.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*, 843–855.

Harris, P. L. (1999). Individual differences in understanding emotion: The role of attachment status and psychological discourse. *Attachment and Human Development*, *1*, 307–324.

Harris, P. L., Olthof, T., Terwogt, M. M., & Hardman, C. E. (1987). Children's knowledge of the situations that provoke emotion. *International Journal of Behavioral Development*, *10*, 319–344.

Ireton, H. (1992). *Child development inventory*. Minneapolis, MN: Behavioral Science Systems.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *21*, 18–23.

Johnston, J., & Wong, A. (2002). Cultural differences in beliefs and practices concerning talk to children. *Journal of Speech, Language, and Hearing Research*, *45*, 916–926.

Kagitcibasi, C. (2005). Autonomy and relatedness in cultural context: Implications for self and family. *Journal of Cross-Cultural Psychology*, *36*, 403–422.

Kim, M. S. (2002). *Non-Western perspectives on human communication: Implications for theory and practice*. Thousand Oaks, CA: Sage.

Lagattuta, K., & Wellman, H. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking and emotion. *Child Development*, *72*, 82–102.

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. M. (1996). Social influences on false belief access: Specific sibling influences or general apprenticeship? *Child Development*, *67*, 2930–2947.

Li, G. (2006). What do parents think? Middle-class Chinese immigrant parents' perspectives on literacy learning, homework, and school-home communication. *The School Community Journal*, *16*(2), 25–44.

Lillard, A. S. (1998). Theories behind theories of mind. *Human Development*, *41*, 40–44.

Lu, H., Su, Y., & Wang, Q. (2008). Talking about others facilitates theory of mind in Chinese preschoolers. *Developmental Psychology*, *44*, 1726–1736.

Maas, A., Kurasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures: A cross-linguistic analysis of person representation. *Journal of Personality and Social Psychology*, *90*, 734–750.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

Martini, M. (1998). Family dynamics during mealtimes in two cultural groups in Hawaii. *Family Development and Parenting*, *5*, 23–34.

Matsumoto, D., Kudoh, T., Scherer, K. R., & Wallbott, H. (1988). Emotional antecedents and reactions in the U.S. and Japan. *Journal of Cross-Cultural Psychology*, *19*, 267–286.

Meins, E., & Fernyhough, C. (1999). Linguistic acquisitional style and mentalising development: The role of maternal mind-mindedness. *Cognitive Development*, *14*, 363–380.

Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Das Gupta, M., Fradley, E., et al. (2003). Pathways to understanding mind: Construct validity and predictive validity of maternal mind-mindedness. *Child Development*, *74*, 1194–1211.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of Theory of Mind understanding. *Child Development*, *73*, 1715–1726.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*, 961–978.

Naito, M., & Koyama, K. (2006). The development of false-belief understanding in Japanese children: Delay and difference? *International Society for the Study of Behavioral Development*, *30*, 290–304.

Ornstein, P. A., Haden, C. A., & Hedrick, A. M. (2004). Learning to remember: Social-communicative exchanges and the development of children's memory skills. *Developmental Review*, *24*, 374–395.

Pons, F., Harris, P., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organizations. *European Journal of Developmental Psychology*, *1*, 127–152.

Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, *8*, 403–430.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14 and 18-month-olds. *Developmental Psychology*, *33*, 12–21.

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, *34*, 161–174.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relationship between children's and mother's mental state language and theory of mind understanding. *Child Development*, *73*, 734–751.

Russell, J., & Bullock, M. (1986). One the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology*, *22*, 97–102.

Scherer, K. R., Matsumoto, D., Wallbott, H., & Kudoh, T. (Eds.). (1988). *Emotional experience in cultural context: A comparison between Europe, Japan and the U.S.A.* Hillsdale, NJ: Erlbaum.

Shweder, R. A., Goodnow, J., Hatano, G., LeVine, R. A., Markus, H., & Miller, P. (1998). The cultural psychology of development: One mind, many mentalities. In R. M. Lerner (Vol. Ed.) & W. Damon (Series Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 865–923). New York: Wiley.

Slaughter, V., Peterson, C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum. *Child Development*, *78*, 839–858.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312), Washington, DC: American Sociological Association.

Stevenson, H. W., Chen, C., & Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese and American children: Ten years later. *Science*, *259*, 53–58.

Symons, D. (2004). Mental state discourse, theory of mind, and the internalization of self-other understanding. *Developmental Review*, *24*, 159–188.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speaker's early vocabularies. *Developmental Psychology*, *32*, 492–504.

Tardif, T., & Wellman, H. (2000). Acquisition of mental state language in Mandarin and Cantonese speaking children. *Developmental Psychology*, *36*, 25–43.

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relate to desire language and emotion understanding. *Child Development*, *77*, 465–481.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development*, *79*, 284–302.

Wang, Q. (2001). "Did you have fun?" American and Chinese mother–child conversations about shared emotional experiences. *Cognitive Development*, *16*, 693–715.

Wang, Q. (2003). Emotion situation knowledge in American and Chinese preschool children and adults. *Cognition and Emotion*, *17*, 725–746.

Wang, Q. (2004). The emergence of cultural self-construct: Autobiographical memory and self-description in American and Chinese children. *Developmental Psychology*, *40*, 3–15.

Wang, Q. (2006). Relations of maternal style and child self-concept to autobiographical memories in Chinese, Chinese immigrant, and European American 3-year-olds. *Child Development*, *77*, 1794–1809.

Wang, Q. (2008). Emotion knowledge and autobiographical memory across the preschool years: A cross-cultural longitudinal investigation. *Cognition*, *108*, 117–135.

Wang, Q., Hutt, R., Kulkofsky, S., McDermott, M., & Wei, R. (2006). Emotion situation knowledge and autobiographical memory in Chinese, immigrant Chinese, and European American 3-year-olds. *Journal of Cognition and Development*, *7*, 95–118.

Wang, Q., & Leichtman, M. D. (2000). Same beginnings, different stories: A comparison of American and Chinese children's narratives. *Child Development*, *71*, 1329–1346.

Wang, Q., Leichtman, M. D., & Davies, K. (2000). Sharing memories and telling stories: American and Chinese mothers and their three year olds. *Memory*, *8*, 159–177.

Wellman, H., Harris, P. L., Banerjee, M., & Sinclair, A. (1995). Early understanding of emotion: Evidence from natural language. *Cognition and Emotion*, *9*, 117–149.

Wellman, H., Phillips, A., & Rodriguez, T. (2000). Young children's understanding of perception, desire and emotion. *Child Development*, *71*, 895–912.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1504–1519

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance

Dante Cicchetti
University of Minnesota

Fred A. Rogosch
University of Rochester

Mark L. Howe
Lancaster University

Sheree L. Toth
University of Rochester

This investigation examined basic memory processes, cortisol, and dissociation in maltreated children. School-aged children (age range = 6–13), 143 maltreated and 174 nonmaltreated, were administered the California Verbal Learning Test–Children (D. C. Delis, J. H. Kramer, E. Kaplan, & B. A. Ober, 1994) in a week-long camp setting, daily morning cortisol levels were assessed throughout the duration of camp, and behavioral symptoms were evaluated. Maltreatment and cortisol regulation were not related to short- or long-delay recall or recognition memory. However, children experiencing neglect and/or emotional maltreatment and low cortisol evinced heightened false recognition memory. Dissociative symptoms were higher in maltreated children; however, high dissociation was related to recognition inaccuracy only among nonmaltreated children. Results highlight the interplay between maltreatment and hypocortisolism in children's recognition memory errors.

Do maltreated children's basic memory processes operate in ways that are fundamentally different from those of nonmaltreated children? Despite the importance that answers to this question hold for advancing research, clinical, and sociolegal issues in the field of child maltreatment (Howe, 2000; Howe, Goodman, & Cicchetti, 2008; Howe, Toth, & Cicchetti, 2006; Toth & Cicchetti, 1998), there has been a paucity of studies, most of which have been published in recent years, that have examined basic memory processes in maltreated children (Howe, Cicchetti, & Toth, 2006).

Child Maltreatment and Memory

There are a number of reasons why the experience of child maltreatment might be expected to alter basic memory processes (Howe, Toth, et al., 2006). Child abuse and neglect are stressful and threatening experiences that may bring about physiological changes in the neural mechanisms related to memory storage (Bremner & Narayan, 1998; Cicchetti, Rogosch, Gunnar, & Toth, 2010; Howe, Cicchetti, et al., 2006). Moreover, a number of the socioemotional sequelae of child maltreatment, including the development of insecure, often disorganized, attachment relationships and deviations in symbolic and representational abilities (Cicchetti & Toth, 2005), may affect the processing of memories in abused and neglected children (Lynch & Cicchetti, 1998; Pollak, Cicchetti, & Klorman, 1998; Rieder & Cicchetti, 1989). In addition, it is conceivable that failures in caregiving may lead maltreated children either to defensively distort memories or to dissociate their memories of their abusive and neglectful experiences (Cicchetti & Valentino, 2006). Finally, maltreated children are at heightened risk for the development of trauma-related psychopathology, including posttraumatic stress disorder (PTSD) and dissociation, that may contribute to deficits in memory performance (Cicchetti & Toth, 2005; Howe, Toth, et al., 2006).

The majority of the investigations on the effects of maltreatment experiences on memory have focused on examinations of adults who retrospectively report having been abused during their childhoods, adults with PTSD, nonmaltreated

---

This research was supported by funding from the National Institute on Drug Abuse (R01 DA17741) and the Spunk Fund Inc.

Correspondence concerning this article should be addressed to Dante Cicchetti, Institute of Child Development, University of Minnesota, 51 East River Road, Minneapolis, MN 55455. Electronic mail may be sent to cicchett@umn.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0014

children participating in analog experiments, or individuals with relatively acute and/or routine stressors involving treatment for accidental injuries or visits to physicians' offices (Howe, 2000; Toth & Cicchetti, 1998). Among the studies of memory in adults who report having been maltreated, several investigations have used the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995) to examine false memories in adults who report having been maltreated in childhood and in adults who did not experience maltreatment (for a review, see Howe, Toth, et al., 2006). Each of these investigations found that adults who report histories of maltreatment evinced a higher frequency of false recognition memory errors (i.e., inaccurately recognizing words not previously presented as heard) than did nonmaltreated comparison individuals.

Despite the fact that there have been relatively few studies of basic memory processes in maltreated children conducted to date, a review of the extant literature reveals that the experience of child abuse and neglect does not appear to exert any special effects on memory that fundamentally change the operation of memory from that observed in nonmaltreated children (Howe, Toth, et al., 2006). For example, no differences have been found in investigations of maltreated children's memory for eyewitnessed events or in their degree of susceptibility to misinformation and suggestion, unless accompanying comorbid psychopathology is present (Eisen, Goodman, Qin, Davis, & Crayton, 2007; Eisen, Qin, Goodman, & Davis, 2002; Howe, Toth, et al., 2006).

In an investigation of true and false recall memory in maltreated children, Howe, Cicchetti, Toth, and Cerrito (2004) examined the memory performances of maltreated and nonmaltreated children on the DRM task. The same pattern of results was found regardless of maltreatment status. Consistent with the normative literature on false memory (Brainerd & Reyna, 2005), both maltreated and nonmaltreated children exhibited false recognition memories and memory inaccuracies with age at the same rate. Similarly, Porter, Lawson, and Bigler (2005) found no significant differences in memory function between sexually abused and nonabused children, despite elevations in psychopathology and diminished performance on measures of attention and executive function for the abused group. Additionally, Beers and DeBellis (2002) did not find any between group differences in basic memory processes for maltreated children and adolescents with PTSD and normative comparison youth. Overall, because these investigations focused on group differences, studies that examine within-group differences might yield different results. Not all maltreated individuals are equally affected by their traumatic experiences (Cicchetti & Toth, 2005). Thus, investigations of basic memory processes in children who have experienced sexual and physical abuse, neglect, and emotional maltreatment are warranted.

Toward this end, a depth-of-processing incidental recall task was used to examine self-relevant information among maltreated children and the relation between maltreatment and false recall on dissociation (Valentino, Cicchetti, Rogosch, & Toth, 2008b). Abused, neglected, and nonmaltreated children all demonstrated better recall for words when words were encoded under self-referent conditions rather than under structural conditions. These results suggest that self-representations serve as an organizing schema for memory and may be used to facilitate recall (cf. Howe, 1998, 2000). Consistent with prior research (Howe et al., 2004), differences were not found among abused, neglected, or nonmaltreated children in the occurrence or total amount of false recall errors. Maltreated children were not rendered more or less susceptible to false recall errors (cf. Howe et al., 2004).

Notably, however, the younger group of neglected children demonstrated a greater proportion of negative false recall errors and less positive recall than did the younger abused children, suggesting a greater vulnerability for memory inaccuracy among neglected children. Among the older group, both abused and neglected children demonstrated less positive false recall than did the nonmaltreated children. Negative self-schemas were associated with increased dissociation among the abused children, whereas positive self-schemas were related to increased dissociation for the neglected children. In addition to addressing the role of dissociation on memory, it also is important to attend to possible neurobiological influences.

HPA-Axis Functioning and Memory

Child abuse and neglect are stressful and threatening experiences that create adaptational challenges (Cicchetti & Rogosch, 2001a; DeBellis, 2001; Hart, Gunnar, & Cicchetti, 1996). One of the physiological systems that has evolved in mammals to help direct and sustain cognitive, emotional, and behavioral activity in response to threat and stress is the hypothalamic–pituitary–adrenal (HPA) axis. Basal levels of cortisol, usually referred to as the "stress hormone," are necessary for normal neurobiological growth and for the support of the metabolic activity necessary to sustain general functioning (McEwen, 1998). Among its many influences, cortisol affects the central neural processes that are implicated in cognition, emotion, and memory (Gunnar & Vazquez, 2006).

Brief elevations in cortisol following acute stressors enhance the individual's ability to manage stressful experiences competently, both physiologically and behaviorally (Lopez, Akil, & Watson, 1999; Vazquez, 1998). However, chronic hyperactivity of the HPA axis may eventuate in a number of negative biological and psychological outcomes, including the loss of hippocampal neurons and impaired affective and cognitive functioning (Sapolsky, 1992, 1996). Although studies of the effects of stress on HPA functioning in humans are less well controlled than are those conducted with animals, hippocampal atrophy is associated with prolonged exposure to glucocorticoids, a class of steroid hormone that includes cortisol. Furthermore, chronic exposure to elevated levels of glucocorticoids is accompanied by cognitive problems, including impairment of declarative memory and visual episodic memory (for reviews, see Heffelfinger & Newcomer, 2001; Howe, Toth, et al., 2006).

Additionally, there is a phenomenon known as hypocortisolism in which individuals who have experienced chronic stress, such as ongoing maltreatment, may exhibit reduced adrenocortical secretion, reduced adrenocortical reactivity, or enhanced negative feedback inhibition of the HPA axis (Gunnar & Vazquez, 2001; Heim, Ehlert, & Hellhammer, 2000; Miller, Chen, & Zhou, 2007). This flattening of the diurnal decrease in cortisol predicts heightened risk of impairment in cardiovascular and immune system functioning (Heim, Newport, Mletzko, Miller, & Nemeroff, 2008). Since it is firmly established that the glucocorticoid cortisol affects the aspects of brain function such as visual episodic memory, just as is the case with hypercortisolism, hypocortisolism also may exert negative impacts upon brain development and cognitive functioning. To the best of our knowledge, to date there have been no studies conducted on the association between memory performance and hypocortisolism, either in maltreated or nonmaltreated children.

Studies cohere in the conclusion that the HPA axis of maltreated children is often characterized by regulatory difficulties (Tarullo & Gunnar, 2006). Although studies have not revealed differences in cortisol regulation between maltreated and nonmaltreated children, differences have emerged when maltreatment subtype is examined. For example, children who have been extensively maltreated (i.e., experienced sexual and physical abuse, plus neglect and/or emotional abuse) have been found to exhibit substantial elevations in morning cortisol levels (Cicchetti & Rogosch, 2001a); children in the multiple abuse (sexual and physical) group also have been shown to display elevations in cortisol levels in both the morning and afternoon (Cicchetti & Rogosch, 2001a). In contrast, maltreated children who have been physically abused, in the absence of sexual abuse, have been shown to exhibit low levels of cortisol. Thus, not all maltreated children exhibit the same patterns of cortisol regulation (Cicchetti, 2002). Hypercortisolism and hypocortisolism are likely outcomes of HPA-axis dysregulation, while some maltreated children display what appears to be normal cortisol regulation (Cicchetti & Rogosch, 2001a; Tarullo & Gunnar, 2006).

Animal models of the impact of early caregiving adversity have predominantly focused on increased reactivity of the HPA axis on outcome (Heim, Owen, Plotsky, & Nemeroff, 1997; Sanchez, Ladd, & Plotsky, 2001; Sanchez et al., 2005). However, as the research has been extended to human and nonhuman primates, it has become increasingly clear that a more accurate prediction is that the HPA axis will likely exhibit evidence of *dysregulation* and that this dysregulation may involve both hyper- and hypofunctioning (Heim, Plotsky, & Nemeroff, 2004; McEwen, 2000). Although not yet fully understood, a number of factors have been shown to affect whether elevated or suppressed cortisol occurs. The HPA axis tends to dysregulate in response to prolonged hyperactivation; thus, in response to chronic stressors, first hyper- and later hypoactivity of the axis may be noted (for a review, see Miller et al., 2007). Accordingly, hypercortisolism is more likely to occur as a result of acute stress and a shorter time since the onset of stress or maltreatment, whereas hypocortisolism is likely the result of a long-term process of chronic stress and a later time since the onset of stress or maltreatment.

Very few studies have examined stress and memory in normative samples of children (Bremner & Narayan, 1998; Bremner & Vermetten, 2001; Howe, Toth, et al., 2006). The only investigation that has examined cortisol and memory in maltreated children conducted to date is that of Eisen et al. (2007). In the Eisen et al. study, maltreated and nonmaltreated comparison children's memory and suggestibility were assessed for an invasive medical procedure that had been administered to them as part of a child abuse and neglect investigation. At the conclusion of the medical examination, older (as compared to younger) children and those who possessed higher cognitive functioning exhibited fewer memory inaccuracies and less suggestibility in their responses to experimenter questions that focused on the recollection of events that were associated with the medical procedures that had been administered. For children low in dissociative symptomatology, higher levels of cortisol reactivity, assessed approximately 20 min after the medical procedures, were found to be related to fewer errors of omission to specific questions about the medical examination and to fewer errors of commission to the experimenter's misleading questions.

In contrast, for children who reported high levels of dissociative symptoms, higher levels of cortisol reactivity were related to more errors of omission to specific questions about the medical procedure and to more errors of commission to misleading questions of the experimenter. Although the initial findings of the Eisen et al. (2007) study are suggestive, limitations in the method of cortisol sampling, lack of cortisol collection for a sizable portion of the sample, and the very specific episodic memory assessment indicate that further research on cortisol, memory, and dissociation in maltreated children is necessary.

An interesting finding that emerged was that children who had experienced sexual and/or physical abuse evinced greater memory accuracy than did those who had experienced child neglect. Specifically, the sexually and/or physically abused children were less suggestible than were the neglected children when confronted with misleading questions about the medical examination, exhibited fewer errors of omission to questions about abuse-related incidents, and provided more correct answers in response to open-ended questions about the medical procedures. It may be the case that abuse contributes to greater vigilance and monitoring of experience and consequently to greater precision in memory encoding and retrieval.

In summary, because child maltreatment experiences do not affect all abused and neglected individuals in the same fashion (known as multifinality; see Cicchetti & Rogosch, 1996), there are likely to be within group differences in memory performance among children with varying maltreatment subtypes. The chronic stress of maltreatment suggests that the cortisol dysregulation characteristic of many maltreated children may be a potential risk for recognition memory inaccuracies in these youngsters. Furthermore, neglected children may be more susceptible to recognition memory errors than are abused children (Eisen et al., 2007; Valentino et al., 2008b). Finally, dissociation appears to be either related to, or an independent contributor to, recognition memory inaccuracies.

In the present article, we report the results of an investigation that examines the relation among neuroendocrine regulation, memory, and dissociative symptoms in maltreated children. We expect the memory performance of maltreated children to differ in relation to varying levels of cortisol regulation. Specifically, we examine the extent to which maltreated children's basic recognition memory processes differ from those of a group of demographically comparable nonmaltreated children. We then investigate the relation between these basic memory processes and individual differences in neuroendocrine functioning and dissociation in maltreated and nonmaltreated children.

This research is guided by the following hypotheses and research questions:

1. In accord with the extant literature, we expected that maltreated children's basic recall and recognition memory skills would not differ from those of nonmaltreated children.
2. We did not expect to find group mean differences between maltreated and comparison children in average morning cortisol levels; however, more extreme patterns of atypical cortisol regulation were expected in physically and/or sexually abused children.
3. We expected that maltreated children with either low morning cortisol or high morning cortisol, suggesting neuroendocrine dysregulation, would have more false recognition memory than nonmaltreated children. Based on the Eisen et al. (2007) and Valentino et al. (2008b) findings, we hypothesized that this relation would be particularly true for children who experienced physical neglect and/or emotional maltreatment but not physical and sexual abuse.
4. Finally, we expected that high levels of dissociation would be related to anticipated differences in memory performance based on maltreatment subtype and cortisol regulation patterns.

Method

Participants

The participants in this investigation included 317 children who attended a recreational summer camp program designed for school-aged low-income, maltreated ($n = 143$) and nonmaltreated ($n = 174$) children. Two cohorts of children from different camp years were aggregated for the current study. In addition to recreational activities, the annual summer camp program incorporated research assessments. The children on average were 9.17 years of age ($SD = 2.43$, range = 6–13), and 54.6% were boys. The racial/ethnic composition of the sample was diverse: 56.6% were Black, 21.5% were White, 17.0% were Latino, and 1.9% were of other racial/ethnic backgrounds.

Parents of all maltreated and nonmaltreated children provided informed consent for their child's participation, as well as consent for examination of any Department of Human Services (DHS) records. Children in the maltreated group had been identified by the county DHS as having experienced child abuse and/or neglect, and the sample was representative of the children in families receiving services from the DHS. A recruitment liaison from DHS contacted eligible maltreating families and explained the study, and if parents were interested their names were released to the project team for recruitment. Comprehensive searches of DHS records were completed, and maltreatment information was coded utilizing operational criteria contained in the Barnett, Manly, and Cicchetti (1993) nosological system for child maltreatment. Children in the maltreated group were living in their family homes and not residing in foster care.

Consistent with national demographic characteristics of maltreating families (NIS–3; Sedlack & Broadhurst, 1996), the maltreated children were predominantly from low-socioeconomic-status families. Accordingly, a demographically comparable sample of nonmaltreated children was recruited from families receiving Temporary Assistance to Needy Families. A DHS recruitment liaison contacted eligible nonmaltreating families and described the project, and if interested, parents signed a release for their names to be given to the project for recruitment. DHS record searches were completed for these families to verify the absence of any record of child maltreatment. Trained research assistants also interviewed mothers of children recruited for the nonmaltreatment group to confirm a lack of DHS involvement and prior maltreatment experiences (Cicchetti, Toth, & Manly, 2003). Subsequently, record searches were conducted in the year following camp attendance to verify that all available information had been accessed. Only children from families without any history of documented or mother-reported abuse and/or neglect were retained in the nonmaltreatment group. In addition, families who had received preventive services through DHS due to concerns over risk for maltreatment also were excluded from the nonmaltreated comparison group to reduce the potential for unidentified maltreatment existing within this group.

Table 1 presents demographic information for families of children in the maltreated and nonmaltreated groups. The groups were comparable on a number of family characteristics. The maltreated and nonmaltreated groups did not differ in terms of total family income, including public assistance, $t(305) = 1.45$, $p = .15$; maternal marital status, $\chi^2(2) = 4.95$, $p = .08$; or maternal education, $\chi^2(1) = 3.11$, $p = .08$. Overall, the majority of families in both groups were headed by a single parent, and most mothers had a high school education or less. Family income was limited, and across the families, 73% were currently receiving public assistance.

Procedure

Children attended a week-long day camp program, participated in research assessments, and took part in a variety of recreational activities. The camp was not designed to provide a therapeutic intervention experience. After providing assent, children participated in various research assessments (see Cicchetti & Manly, 1990, for detailed descriptions of camp procedures). Children were administered the specific measures of interest in the current report, the California Verbal Learning

Table 1
Family Demographic Features

	Nonmaltreated	Maltreated
	M (*SD*) or %	*M* (*SD*) or %
Marital status		
Never married	35.7%	28.2%
Married/living together	42.1%	38.7%
No longer married	22.2%	33.1%
Maternal education		
High school or less	52.0%	62.0%
Total family income		
$1,000s including public assistance	24.81 (12.37)	22.86 (10.90)

Test–Children (CVLT–C) and the Peabody Picture Vocabulary Test–III (PPVT–III), on different days during the camp week. At the end of each week, counselors who worked with the children rated them on behavioral symptomatology.

Measures

Maltreatment classification system. The Maltreatment Classification System (MCS; Barnett et al., 1993) was used to delineate diverse features of maltreatment experienced by individual children. The MCS utilizes DHS records detailing investigations and findings involving maltreatment in identified families over time. Rather than relying on official designations and case dispositions, the MCS codes all available information from DHS records, making independent determinations of maltreatment experiences. Based on operational criteria, the MCS designates all of the subtypes of maltreatment children have experienced (i.e., neglect, emotional maltreatment, physical abuse, sexual abuse), the perpetrator(s) of the maltreatment, the severity of each subtype, frequency of occurrence, and the developmental periods of occurrence, allowing for determination of onset, recency, and chronicity.

In terms of the subtypes of maltreatment, *neglect* involves failure to provide for the child's basic physical needs for adequate food, clothing, shelter, and medical treatment. In addition to inadequate attention to physical needs, forms of this subtype include lack of supervision, moral-legal neglect, and educational neglect. *Emotional maltreatment* involves extreme thwarting of children's basic emotional needs for psychological safety and security, acceptance and self-esteem, and age-appropriate autonomy. Examples of emotional maltreatment of increasing severity include belittling and ridiculing the child, extreme negativity and hostility, exposure to severe marital violence, abandoning the child, and suicidal or homicidal threats. *Physical abuse* involves the nonaccidental infliction of physical injury on the child (e.g., bruises, welts, burns, choking, broken bones). Injuries range from minor and temporary to permanently disfiguring. Finally, *sexual abuse* involves attempted or actual sexual contact between the child and caregiver for purposes of the caregiver's sexual satisfaction or financial benefit. Events range from exposure to pornography or adult sexual activity, to sexual touching and fondling, to forced intercourse with the child.

Coding of the DHS records was conducted by trained research assistants, doctoral students, and clinical psychologists. Coders were required to meet acceptable reliability with criterion standards before coding actual records for the study. Coders demonstrated acceptable reliability (weighted κs with the criterion = .86 to .98). Reliabilities for the presence versus absence of maltreatment subtypes (κs = .90–1.00) also were established. Other investigators have demonstrated that the MCS is reliable and valid in classifying maltreatment (Bolger, Patterson, & Kupersmidt, 1998; Dubowitz et al., 2005; English et al., 2005; Manly, 2005; Smith & Thornberry, 1995).

Among the maltreated children, 79.7% had experienced neglect, 73.4% had experienced emotional maltreatment, 36.4% had experienced physical abuse, and 14.7% had experienced sexual abuse. Thus, neglect and emotional maltreatment had occurred frequently among most of the maltreated children, whereas lower prevalence rates were found for physical and sexual abuse. As is typical in maltreated populations (Bolger et al., 1998; Manly, Cicchetti, & Barnett, 1994; Manly, Kim, Rogosch, & Cicchetti, 2001), the majority of children had experienced multiple subtypes of maltreatment. Specifically, 68.5% of the maltreated children had experienced two or more maltreatment subtypes ($M = 2.04$, $SD = 0.89$), and 14 of the 15 possible combinations of the four maltreatment subtypes were observed in the sample. In order to consolidate groups for comparison purposes, we sought to differentiate children whose maltreatment experiences had involved physical and/or sexual abuse with or without neglect and emotional maltreatment (PA/SA; $n = 61$) from those who had not been physically or sexually abused but had been neglected and/or emotionally maltreated (PN/EM; $n = 82$). Within the PA/SA group, 75.4% had been emotionally maltreated and 82.0% had been neglected. Comparable rates of emotional maltreatment (72.0%) and neglect (78.0%) were observed in the PN/EM group. Thus, what distinguished the PA/SA group from the PN/EM group was the presence of abuse.

Salivary cortisol. The camp context provided a setting where a consistent collection of saliva samples could occur at uniform times across the camp week. Each day upon arrival to the camp at 9:00 a.m. trained research assistants obtained saliva samples from each child. The children had not consumed food or drink for at least 30 min before each saliva sample was obtained. Given the approximate 45-min transportation time and the time spent being greeted by camp staff, children had been awake at least 1 hr prior to providing the morning saliva samples. Because cortisol levels are volatile and sharply rise and fall during the first hour after morning awakening (cf. Susman et al., 2007), the timing of morning samples likely occurred when relative stability in morning level was established, before gradually declining over the course of the day. Although we were not able to obtain daily wake-up times for children, there is no reason to suspect that wakening times systematically varied across groups.

Two different saliva collection procedures were used depending on the child's cohort membership. In the first cohort involving 49.8% of the participants, Kool-Aid crystals were administered to stimulate saliva flow, mouthing of a sterile cotton dental role to soak up the saliva, with subsequent expressing of the saliva from the dental role into storage vial via a syringe at −80 °C. Cortisol assays were conducted at the University of Minnesota utilizing the Ciba Corning Magic Cortisol Kit. For the second cohort, children deposited saliva through a small straw directly into storage vials. The samples were analyzed at Pennsylvania State University using a high-sensitive enzyme immunoassay (Salimetrics, PA).

Because of the difference in assay method and variation in the distribution of assay results in different cohorts, standardized scores were used to calibrate samples across cohorts. As is typical, high kurtosis and negative skew characterized the distributions of cortisol values each year. As a result, all values were transformed using log10. Within each year, all log10-transformed values were converted to Z scores. The standardized scores thus allowed for relative comparability in cortisol levels across camp years (cf. Cicchetti & Rogosch, 2001a, 2001b). All morning values for each child were averaged to derive a mean morning cortisol variable. The two cohorts did not differ on demographic characteristics, cortisol levels, memory variables, or dissociation.

California Verbal Learning Test–Children. The CVLT–C (Delis, Kramer, Kaplan, & Ober, 1994) provides an assessment of strategies and processes involved in learning and recall of verbal material. The test is individually administered and normed for children aged 5–16. It is used to assist in the diagnosis and treatment of memory impairments. The CVLT–C provides the indices of verbal learning and recall and recognition memory. The CVLT–C was developed and standardized on a representative sample of the U.S. population, and reliability and validity have been established (Delis et al., 1994). Completion time for the CVLT–C was approximately 50 min.

In the first five trials, children are read a list of 15 words from three nonaffective semantic categories (i.e., things to wear, things to play with, fruits) and then are asked to recall as many words as they can. An interference trial of 15 new words is then presented. Children are then requested to freely recall the words from the original list and then are given a subsequent trial in which semantic cues to the words are presented to facilitate recall. A 20-min delay involving a nonverbal activity occurred next. Children then are directed to recall the original word list, followed by another cued recall trial. Finally, children were presented with a serial list of 45 items and asked to determine whether each word was on the original list. In the current report, children are compared on their abilities in short-delay recall, long-delay recall, recognition memory, and false positive memory recognition. A discriminability index, comparable to d', which assesses children ability to distinguish target words from distractor words, also was evaluated. Age standardized scores for all memory variables were used. Finally, deficits in retrieval (low standard scores for long-delay recall and normal standard scores for discriminability) and encoding (low standards scores for both long-delay recall and discriminability) also were examined.

Peabody Picture Vocabulary Test–III. The PPVT–III (Dunn & Dunn, 1997) assesses children's receptive vocabulary abilities. The PPVT–III correlates with the Full Scale IQ of the Wechsler Intelligence Test for Children (WISC–III). The PPVT–III was completed in approximately 15 min.

Teacher Report Form and Dissociation subscale. Behavioral symptomatology was evaluated at the end of each week by camp counselors' completion of the Teacher Report Form (TRF; Achenbach, 1991) after approximately 35 hr of observation and interaction with the children. The camp counselors were unaware of maltreatment status and research hypotheses. The TRF is a widely used and validated instrument to assess behavioral disturbance from the perspective of teachers, and the measure was used in the present study, because camp counselors are able to observe similar behaviors to that of teachers. The TRF, containing 118 items rated for frequency, assesses a broad range of symptoms. In the present study, interrater reliability for total behavior problem scores, based on average intraclass correlations among pairs of raters, ranged from .78 to .87 (M = 0.82). Ogawa, Sroufe, Weinfield, Carlson, and Egeland (1997) developed a subscale to assess dissociative symptoms in childhood and adolescence based on 12 TRF items. The content of these items is similar to those assessed by the *Child Dissociative Checklist* (Putnam, Helmers, & Trickett, 1993). The counselors' scores for the dissociation subscale items for each child were averaged to derive individual child scores for dissociative symptomatology.

Results

Maltreatment and Cortisol Dysregulation

Overall, as found in previous work (Cicchetti & Rogosch, 2001a, 2001b), maltreated and nonmaltreated children did not differ in their average level of morning cortisol, $t(313) = 1.60$, $p = .11$. Furthermore, no significant differences were found for the PA/SA, PN/EM, and nonmaltreated groups, $F(2, 312) = 1.41$, $p = .24$ (see Table 2). However, based on an interest in patterns of cortisol dysregulation (i.e., hypocortisolism, hypercortisolism) rather than solely on group mean levels, children were differentiated into three groups based on the sample's distribution of average morning cortisol scores: low morning cortisol (≤ -1 *SD*), medium morning cortisol (> -1 *SD* and $< +1$ *SD*), and high morning cortisol ($\geq +1$ *SD*). The proportion of children in the three cortisol groups differed between maltreated and nonmaltreated children, $\chi^2(2, N = 315) = 6.59$, $p = .04$. Group contrasts revealed that more maltreated children (19.7%) were in the low cortisol group compared to nonmaltreated children (10.4%), $\chi^2(1, N = 315) = 5.43$, $p = .02$, whereas the proportion in the high morning cortisol group did not differ between maltreated (14.8%) and nonmaltreated children (12.1%), $\chi^2(1, N = 315) = 0.47$, $p = .49$.

Maltreatment Subtypes, Cortisol Dysregulation, and Memory Performance

Next, a series of analyses of covariance (ANCOVAs) was conducted to examine maltreatment subtype groups and the three-level grouping of morning cortisol in relation to differences in memory processes, as indicated by standard scores from the CVLT–C, including short-delay recall, long-delay recall, recognition memory, false positive recognition memory, and the discriminability index. Subtype group means are presented in Table 2. Children's age and PPVT–III scores were used as covariates in all analyses. Initially, general memory processes, including short- and long-delay recall and recognition memory were examined. ANCOVAs for these variables did not result in any significant main or interactive effects (see Table 3).

In contrast, significant effects were obtained for false positive recognition memory, as shown in Table 3. In the ANCOVA, a significant effect was obtained for cortisol level, $F(2, 302) = 5.13$, $p = .01$; however, this effect was clarified by a significant interaction between maltreatment subtype and cortisol level, $F(4, 302) = 2.45$, $p = .05$. Figure 1 depicts the interactive effect. To probe the interaction effect, separate ANCOVAs were conducted within each of the maltreatment subtype groups to examine the relation of cortisol level to false recognition memory. The effect of cortisol level was not significant in the analysis for nonmaltreated children, $F(2, 167) = 1.38$, $p = .46$, or for the PA/SA group, $F(2, 55) = 1.97$, $p = .24$. However, a significant effect was obtained for the PN/EM group, $F(2, 76) = 9.01$, $p < .001$. Post hoc Tukey tests indicated that within the PN/EM group of children, the low-cortisol group had significantly higher false recognition memory scores than the average- and high-cortisol groups.

Table 2
Morning Cortisol and California Verbal Learning Test–Children (CVLT–C) Standard Scores for Maltreatment Subtype Groups

	Nonmaltreated	PN/EM	PA/SA
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Morning cortisol Z score	0.07 (0.98)	−0.14 (0.82)	−0.05 (1.27)
CVLT–C standard scores			
Short-delay free recall	−0.05 (0.90)	0.12 (0.91)	−0.05 (1.30)
Long-delay free recall	−0.59 (1.07)	−0.77 (1.15)	−0.64 (1.19)
Recognition memory	−0.23 (1.04)	−0.38 (1.29)	−0.20 (1.13)
False positive recognition	−0.07 (1.21)	0.01 (0.92)	0.27 (1.24)
Discriminability index	−0.24 (1.41)	−0.37 (1.16)	−0.55 (1.43)

Note. PN/EM = physical neglect and/or emotional maltreatment only; PA/SA = physical abuse and/or sexual abuse.

Table 3
Analyses of Covariance With Maltreatment Subtype Group and Morning Cortisol Level for California Verbal Learning Test–Children Memory Indices

Source	*df*	*F*	η^2	*p* value
Short-delay free recall standard score				
Age	1	1.22	.004	.27
PPVT–R standard score	1	6.60	.021	.01
Maltreatment subtype group	2	0.56	.004	.57
Level of morning cortisol	2	1.40	.009	.25
Subtype × Cortisol Level	4	0.21	.003	.93
Error	302			
Long-delay free recall standard score				
Age	1	0.43	.001	.51
PPVT–R standard score	1	7.17	.023	.01
Maltreatment subtype group	2	1.91	.013	.15
Level of morning cortisol	2	2.44	.016	.09
Subtype × Cortisol Level	4	0.93	.012	.45
Error	302			
Recognition memory standard score				
Age	1	2.30	.008	.13
PPVT–R standard score	1	0.06	.000	.89
Maltreatment subtype group	2	1.81	.012	.17
Level of morning cortisol	2	0.16	.001	.85
Subtype × Cortisol Level	4	1.52	.020	.20
Error	302			
False positive recognition standard score				
Age	1	0.12	.000	.73
PPVT–R standard score	1	20.50	.064	.000
Maltreatment subtype group	2	1.17	.008	.31
Level of morning cortisol	2	5.13	.033	.01
Subtype × Cortisol Level	4	2.45	.031	.05
Error	302			
Discriminability index standard score				
Age	1	0.65	.002	.42
PPVT–R standard score	1	14.48	.046	.000
Maltreatment subtype group	2	1.96	.013	.14
Level of morning cortisol	2	3.16	.021	.04
Subtype × Cortisol Level	4	2.28	.029	.06
Error	302			

Note. PPVT–R = Peabody Picture Vocabulary Test–Revised.

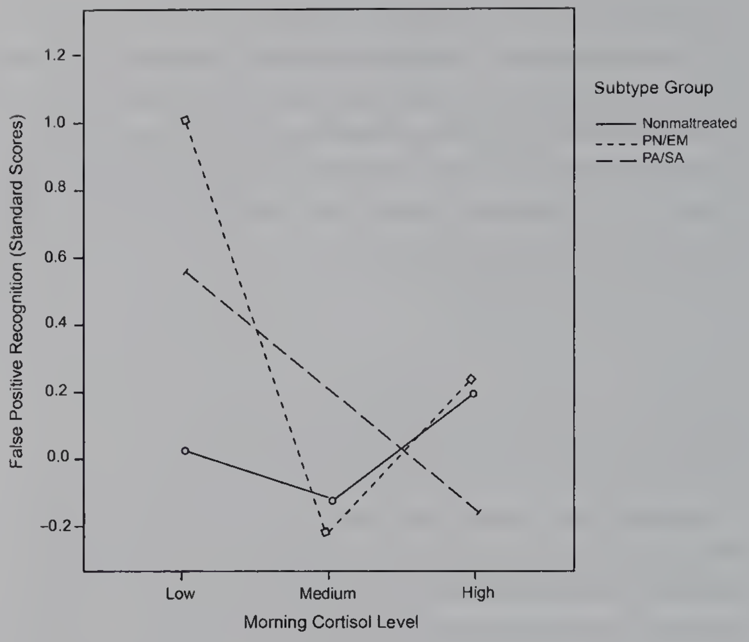


Figure 1. Interaction of maltreatment subtype group and level of morning cortisol for California Verbal Learning Test–Children false recognition memory.
Note. PN/EM = physical neglect and/or emotional maltreatment only; PA/SA = physical abuse and/or sexual abuse.

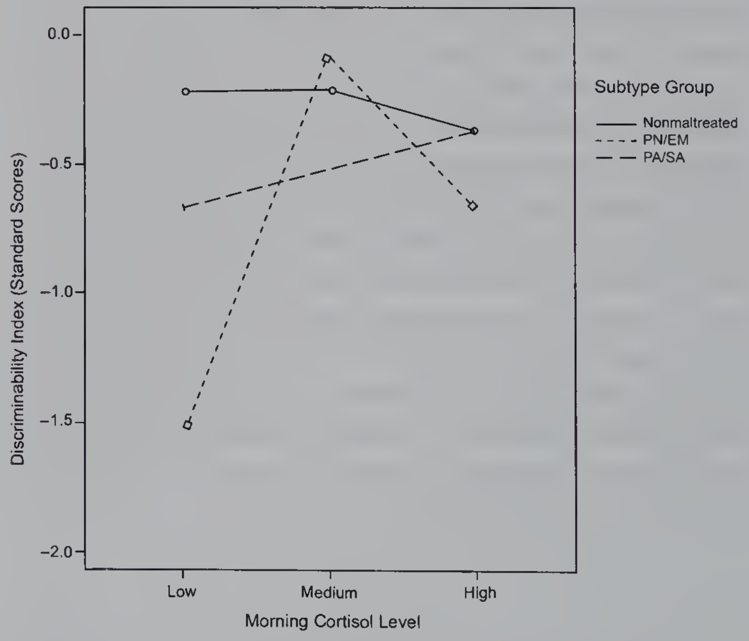


Figure 2. Interaction of maltreatment subtype group and level of morning cortisol for the California Verbal Learning Test–Children discriminability index.
Note. PN/EM = physical neglect and/or emotional maltreatment only; PA/SA = physical abuse and/or sexual abuse.

The discriminability index, which measures a child's ability to distinguish between target words and distractors, taking into account a child's recognition memory ability in relation to the rate of false positives, also was evaluated. False recognition memory and the discriminability index were correlated ($r = .49$, $p < .001$). The results from the ANCOVA for the discriminability index mirrored those observed for false recognition memory. The effect of cortisol levels was significant, $F(2, 302) = 3.16$, $p = .04$, and a marginally significant interaction effect between maltreatment subtype group and cortisol level was observed, $F(4, 302) = 2.28$, $p = .06$. This interaction is depicted in Figure 2. The interaction effect was probed by examining the effect of cortisol level on the discriminability index within each of the maltreatment subtype groups. No significant effect of cortisol level was found in the nonmaltreated group, $F(2, 167) = 0.10$, $p = .91$, or the PA/SA group, $F(2, 55) = 0.14$, $p = .87$, whereas a significant interaction was obtained for the PN/EM group, $F(2, 76) = 11.18$, $p < .001$. Post hoc Tukey tests indicated that among the PN/EM children, those with low morning cortisol group had significantly lower discriminability scores than children with medium level morning cortisol.

Additionally, the potential for retrieval and encoding deficits to influence these findings was evaluated. The overall rates of deficits in retrieval (10.4%) and in encoding (8.2%) in the sample were low. No significant differences between subtype groups were obtained for retrieval deficits, $\chi^2(2, N = 317) = 0.59$, $p = .74$, or for encoding deficits, $\chi^2(2, N = 317) = 1.18$, $p = .55$. Similarly, the cortisol groups also were not differentiated by the rate of retrieval deficits, $\chi^2(2, N = 317) = 0.87$, $p = .65$, or encoding deficits, $\chi^2(2, N = 317) = 0.28$, $p = .87$. Thus, retrieval and encoding deficits did not appear to be unique to particular subtype or cortisol level groups.

The differentiation in memory performance between the PA/SA and PN/EM groups was examined in further detail to determine if other aspects of child maltreatment would better explain the interaction effects of level of morning cortisol and subtype. Specifically, rather than subtype, various maltreatment features, including early onset of maltreatment in infancy and/or toddlerhood, chronicity based on number of developmental periods of maltreatment occurrence, recency of maltreatment, and the number of subtypes of maltreatment experienced were examined in the same ANCOVAs as utilized earlier. None of the interaction effects of the various maltreatment features and level of morning cortisol was significant, indicating that the subtype interactive effects provided the greatest explanatory power.

The Influence of Dissociative Symptomatology

The influence of dissociative symptomatology on the findings also was examined. Initially, an ANOVA was conducted to examine potential differences in dissociation scores based on cortisol levels and subtype group. Neither the main effect of cortisol level, $F(2, 303) = 0.42$, $p = .66$, nor the interaction of subtype group and cortisol level, $F(4, 303) = 0.79$, $p = .53$, was significant. However, significant subtype differences were found, $F(2, 303) = 7.37$, $p < .001$. Tukey post hoc tests indicated that children in both the PA/SA group ($M = 3.53$, $SD = 3.05$) and the PN/EM group ($M = 2.93$, $SD = 2.99$) had significantly higher dissociation scores on average than the nonmaltreated group ($M = 1.63$, $SD = 1.94$).

Next, given an interest in the effects of high levels of dissociative symptoms, children with dissociation symptoms > +1 *SD* were categorized into a high-dissociation group ($n = 51$), and all remaining children were categorized into a low-dissociation group ($n = 260$). The dissociation level factor was then used in subsequent ANCOVAs along with subtype group to predict memory performance, with age and PPVT–R standard scores as covariates. The main and interaction effects for these analyses were all nonsignificant for short-term recall, long-term recall, and recognition memory. In contrast, findings emerged for false recognition memory and discriminability. For false recognition memory, a main effect was found for dissociation level, $F(1, 311) = 4.37$, $p = .04$. However, this main effect was qualified by a significant interaction effect of dissociation level and subtype group, $F(2, 311) = 4.16$, $p = .02$. To probe the interaction, follow-up analyses were conducted within each subtype group. Whereas no significant differences in false recognition were observed between children with high and low dissociation levels in the PN/EM, $F(1, 75) = 1.53$, $p = .22$, or the PA/SA, $F(1, 56) = 1.50$, $p = .23$, groups, among the nonmaltreated group, children with high levels of dissociative symptoms had higher false recognition scores ($M = 0.90$, $SD = 1.66$) than those with low levels of dissociation ($M = -0.17$, $SD = 1.11$), $F(1, 168) = 8.91$, $p = .003$. Similar effects were found for discriminability, with a significant effect for the interaction of dissociation level and subtype group, $F(2, 303) = 3.29$, $p = .04$. Again, no differences were found within the PN/EM or PA/SA based on dissociation level, whereas among nonmaltreated children, those with high dissociation had poorer discriminability scores ($M = -1.20$, $SD = 1.89$) than those with low dissociation scores ($M = -0.13$, $SD = 1.31$), $F(1, 168) = 6.27$, $p = .04$. Finally, whether high dissociation was related to greater likelihood of retrieval or encoding deficits was examined; both chi-square tests were nonsignificant, *ps* > .80. Further analyses examining subtype, dissociation level, and cortisol group were not feasible given restrictions in cell sizes. Overall, the findings suggest that dissociation contributes to higher inaccurate memory among nonmaltreated children but does not influence false recognition or discriminability among PN/EM or PA/SA children. Thus, dissociation does not account for the attenuated memory performance of PN/EM children with low cortisol.

Discussion

Consistent with the extant literature, the findings from this investigation reveal that the experience of child maltreatment per se does not appear to adversely affect, nor does it enhance, basic memory for nontraumatic material (Howe, Cicchetti, et al., 2006; Howe, Toth, et al., 2006). Specifically, in accord with our hypotheses, the basic recall and recognition memory abilities of maltreated children did not differ from that of the nonmaltreated comparison children. Likewise, an examination of the rate of deficits in encoding and retrieval memory processes revealed no between group differences. Conversely, when examining HPA-axis functioning and memory, some interesting differences between maltreated and nonmaltreated comparison children do begin to emerge. As anticipated, differences were not found in the average morning basal cortisol of maltreated and nonmaltreated children. Notably, however, when different levels of cortisol regulation were examined, more maltreated children exhibited lower levels of basal cortisol than did nonmaltreated children.

Our hypothesis that low morning cortisol would result in increased false recognition memory in maltreated children was supported; however, our hypothesis that maltreated children who exhibited high morning cortisol levels would display increases in false recognition memory was not confirmed. More specifically, in the present study the expectation that the subgroup of maltreated children with low cortisol who had experienced PN/EM would exhibit increased false recognition memory and poorer discriminability was confirmed. These results are in keeping with those reported in two prior investigations of false memory in maltreated children (Eisen et al., 2007; Valentino et al., 2008b). In both of these investigations, children who had experienced physical neglect and often co-occurring emotional maltreatment evinced increased levels of memory inaccuracy compared to children who experienced physical abuse and/or sexual abuse.

Furthermore, an in-depth examination of additional MCS variables that could have influenced the results obtained included an investigation of the potential contribution of age of onset of maltreatment, the number of developmental periods during which a child experienced maltreatment, the recency of maltreatment, and the distinct number of subtypes of maltreatment that occurred. None of these factors were found to interact significantly with morning levels of cortisol. Thus, the significant interaction effect of physical neglect and/or emotional maltreatment and low levels of morning cortisol resulted in the clearest identification of a subgroup of maltreated children with elevated recognition inaccuracy.

To the best of our knowledge, the finding that low basal levels of cortisol are associated with impairments in false recognition memory in neglected and/or emotionally maltreated children is the first demonstration that a form of cortisol dysregulation known as hypocortisolism can exert negative effects on memory. Conversely, we did not find an association between high basal levels of cortisol and increased false recognition memory. The absence of differences for high basal cortisol may be related to the co-occurrence of psychopathology, specifically, dissociation. The only previous investigation to examine memory processes and cortisol in maltreated children and adolescents (Eisen et al., 2007) found that among children with high levels of dissociative symptomatology, high cortisol reactivity to a stressful medical examination was associated with memory impairment (i.e., increased suggestibility); however, for children with low levels of dissociation, high cortisol reactivity was not found to relate to suggestibility of memory. Thus, it appears that high levels of dissociation moderated the relation between cortisol reactivity and suggestibility. Unfortunately, due to sample size restrictions, the Eisen et al. (2007) study could not examine the possible effects of maltreatment subtype on cortisol reactivity.

In the current investigation, maltreated children in the physically and/or sexually abused and the physically neglected and/or emotionally maltreated groups had greater levels of dissociative symptoms than did the nonmaltreated comparison children. Cortisol level alone or in interaction with subtype did not predict dissociation symptoms. Similar to the results obtained by Eisen et al. (2007), we found that high dissociation was related to higher false recognition memory; however, in contrast, and contrary to our expectations, in our investigation these results held true only for the nonmaltreated comparison group of children with high dissociative symptoms. We did not find any relation between high dissociation and higher false recognition memory for the physically and/or sexually abused or the physically neglected and/or emotionally maltreated groups. It should be kept in mind that the

Eisen et al. study focused on cortisol reactivity (and not basal levels, as we did) and assessed an affectively laden form of memory (suggestibility in response to a medical exam as opposed to inaccuracies in basic recognition memory).

Our findings suggest that there may be two different risk profiles associated with false recognition memory. For the subgroup of nonmaltreated children, increased levels of dissociative symptoms appear to be the contributor to false recognition memory. In contrast, increased dissociative symptomatology was not related to false recognition memory in the physically neglected and/or emotionally maltreated subgroup of children; rather, low levels of basal cortisol were found to be associated with false recognition memory. Among children in the physically and/or sexually abused group, there does not appear to be a risk profile for recognition memory inaccuracy. Specifically, neither increased dissociation, nor cortisol level were associated with false recognition memory for these children.

Because the HPA axis has been demonstrated to down-regulate in response to chronic stressors, the presence of hypocortisolism suggests that HPA-axis dysregulation has been operating for a prolonged period of time. Just as has been found in empirical studies of hypercortisolism and memory (McEwen & Stellar, 1993; Sapolsky, 1996), the HPA-axis dysregulation characteristic of hypocortisolism also may be related to neurobiological sequelae and memory impairments.

Neuroimaging studies conducted with maltreated individuals have revealed that the hippocampus, the major brain structure that subserves memory, evinces structural decreases in volume among adults (Bremner et al., 2003). However, cross-sectional and short-term longitudinal studies of maltreated children and adolescents with PTSD have not found evidence of structural hippocampal impairments (DeBellis, 2001, 2005). Although stress-induced hypercortisolism has often been invoked as a potential neurobiological mechanism underlying the development of hippocampal damage, it is conceivable that hypocortisolism also may be a likely indicator of hippocampal damage given that the HPA axis of individuals with hypercortisolism over time would tend to down-regulate in response to prolonged hyperactivation.

It is instructive to consider why it was the subgroup of physically neglected and/or emotionally maltreated children with low levels of cortisol who exhibited the increased false recognition memory impairments, whereas the physically and/or sexually abused children with low levels of cortisol, who also had levels of neglect and emotional maltreatment comparable to the PN/EM group, did not reveal increases in false recognition memory. It is possible that the more distinct incidents of trauma experienced by physically and/or sexually abused children may contribute to enhanced attention and greater vigilance to environmental stimuli. In turn, this hypervigilance may counteract the negative effects of hypocortisolism upon memory. The lack of predictability associated with when physical and/or sexual abuse occurs may engender hypervigilance to aggression in these children, even when abusive events are not taking place (Pollak, Cicchetti, Klorman, & Brumaghim, 1997; Rieder & Cicchetti, 1989; Teisl & Cicchetti, 2008). The prior experience of anger and trauma may impel physically and sexually abused children to engage in a more careful monitoring of each context they experience (cf. Pollak & Kistler, 2002; Pollak, Vardi, Putzer-Bechner, & Curtin, 2005; Shackman, Shackman, & Pollak, 2007). Moreover, the chronic nature of neglect most likely involves general rather than discrete memories. Although further investigation is necessary, it may be that a memory system organized around overgeneral memories is more prone to impaired memory performance in the context of hypocortisolism.

The fact that maltreated children who have not experienced physical or sexual abuse exhibit memory impairments in conjunction with hypocortisolism is quite interesting. In fact, various subgroups of maltreated children display increases in false recognition and recall memory under mother-referent and self-referent conditions (Valentino et al., 2008a, 2008b). Why have differences not been found in other forms of basic memory? Why has hypercortisolism not resulted in memory difficulties in maltreated children? Alternatively, does hypercortisolism initiate the pathway from hyper- to hypocortisolism to structural neurobiological changes in the hippocampus that alter basic memory processes?

In order to address these challenging questions in a manner that is commensurate with the complexity of the issues noted, future investigations must be conducted that strive to identify other factors that may exert mediating or moderating effects on false recall and recognition memory. It will be important to conduct prospective longitudinal investigations that repeatedly assess HPA-axis functioning and various forms of basic memory, utilizing diverse memory paradigms. The incorporation of molecular genetic techniques into these studies also could help to elucidate, for example, how variants of glucocorticoid receptor genes and corticotrophin receptor hormone genes, and variations in catechol-o-methyl-transferase genes may serve as moderators of the relation among maltreatment, HPA-axis functioning, and memory. Ideally, subgroups of maltreated children who have either hyper- or hypocortisolism should also receive periodic structural and functional neuroimaging assessments. Such multiple levels of analysis research programs have the potential to provide answers to key questions that the field of memory and trauma must address.

The findings of this investigation contribute to the extant literature by highlighting the interplay between neglect and/or emotional maltreatment and hypocortisolism in relation to false recognition memory. In view of the current findings related to subtype of maltreatment, it is important to note that physically neglected and emotionally maltreated children have been accorded less attention than physically and sexually abused children in research conducted to date on basic memory processes in maltreated children. Although the emphasis on investigating memory in physically and sexually abused children is understandable given that these forms of maltreatment are particularly heinous forms of child maltreatment, it was the physically neglected and emotionally maltreated children who exhibited the impairments in false recognition memory. Thus, future research on memory processes in maltreated children should include groups of physically neglected and emotionally maltreated children (cf. Eisen et al., 2007; Valentino et al., 2008a, 2008b).

Similarly, forensic work typically is more focused on physical and sexual abuse cases. These children, even the subgroup with low cortisol, do not appear to have a greater likelihood of false recognition memory; however, the less studied physically neglected and emotionally maltreated children do show evidence of increases in false recognition memory. The findings of the current investigation urge caution in prematurely assuming that the memories of neglected children are less affected than are those of children who have experienced physical or sexual abuse.

It is important to acknowledge that the findings of the current investigation may not generalize to maltreated children who do not come to the attention of child protective agencies. A potential limitation of the current study is the difficulty in being able to identify sufficient numbers of children who were abused but not neglected and/or emotionally maltreated in order to have distinct group comparisons. However, the co-occurrence of these forms of maltreatment is typical in this population, and an exclusive focus on children who were only abused would result in poor generalizability to the larger maltreated population. Nevertheless, it does not seem likely that the memory performance of the atypical group of children who are exclusively physically and/or sexually abused would be markedly different from those who also have had neglect and/or emotional maltreatment experiences.

The current investigation also possesses implications for intervention. It is imperative that children who have been neglected and/or emotionally maltreated be afforded access to services commensurate with those more commonly provided to children who have been physically or sexually abused. Because neglected and/or emotionally maltreated children may evidence perturbations in memory associated with their experiences, the provision of narrative therapies that assist them in developing a cohesive life story in the context of neglect and emotional maltreatment may be particularly useful (Fivush, Bohanek, Robertson, & Duke, 2004). Such therapeutic approaches ultimately may prevent the coalescence of the negative self-representations frequently associated with experiences of maltreatment (Toth, Cicchetti, Macfie, Rogosch, & Maughan, 2000; Toth, Maughan, Manly, Spagnola, & Cicchetti, 2002).

References

Achenbach, T. (1991). *Manual for the Teacher Report Form and 1991 profile*. Burlington: Department of Psychiatry, University of Vermont.

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. L. Toth (Eds.), *Child abuse, child development, and social policy* (pp. 7–73). Norwood, NJ: Ablex.

Beers, S. R., & DeBellis, M. D. (2002). Neuropsychological function in children with maltreatment-related stress disorder. *American Journal of Psychiatry*, *159*, 483–486.

Bolger, K. E., Patterson, C. J., & Kupersmidt, J. B. (1998). Peer relationships and self-esteem among children who have been maltreated. *Child Development*, *69*, 1171–1197.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Bremner, J. D., & Narayan, M. (1998). The effects of stress on memory and the hippocampus throughout the life cycle: Implications for childhood development and aging. *Development and Psychopathology*, *10*, 871–885.

Bremner, J. D., & Vermetten, E. (2001). Stress and development: Behavioral and biological consequences. *Development and Psychopathology*, *13*, 473–489.

Bremner, J. D., Vythilingam, M., Vermetten, E., Southwick, S. M., McGlashan, T., Nazeer, A., et al. (2003). Deficits in hippocampal structure and function in women with childhood sexual abuse-related posttraumatic stress disorder (PTSD) measured with magnetic resonance imaging and postitron emission tomography. *American Journal of Psychiatry*, *160*, 924–932.

Cicchetti, D. (2002). The impact of social experience on neurobiological systems: Illustration from a constructivist view of child maltreatment. *Cognitive Development*, *17*, 1407–1428.

Cicchetti, D., & Manly, J. T. (1990). A personal perspective on conducting research with maltreating families: Problems and solutions. In G. Brody & I. Sigel (Eds.), *Methods of family research: Families at risk* (Vol. 2, pp. 87–133). Hillsdale, NJ: Erlbaum.

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, *8*, 597–600.

Cicchetti, D., & Rogosch, F. A. (2001a). Diverse patterns of neuroendocrine activity in maltreated children. *Development and Psychopathology*, *13*, 677–694.

Cicchetti, D., & Rogosch, F. A. (2001b). The impact of child maltreatment and psychopathology upon neuroendocrine functioning. *Development and Psychopathology*, *13*, 783–804.

Cicchetti, D., Rogosch, F. A., Gunnar, M. R., & Toth, S. L. (2010). The differential impacts of early abuse on internalizing problems and diurnal cortisol activity in school-aged children. *Child Development*, *81*, 252–269.

Cicchetti, D., & Toth, S. L. (2005). Child maltreatment. *Annual Review of Clinical Psychology*, *1*, 409–438.

Cicchetti, D., Toth, S. L., & Manly, J. T. (2003). *Maternal maltreatment interview*. Unpublished manuscript, Rochester, NY.

Cicchetti, D., & Valentino, K. (2006). An ecological transactional perspective on child maltreatment: Failure of the average expectable environment and its influence upon child development. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation* (2nd ed., pp. 129–201). New York: Wiley.

DeBellis, M. D. (2001). Developmental traumatology: The psychobiological development of maltreated children and its implications for research, treatment, and policy. *Development and Psychopathology*, *13*, 539–564.

DeBellis, M. D. (2005). The psychobiology of neglect. *Child Maltreatment*, *10*, 150–172.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *CVLT–C manual*. San Antonio, TX: Psychological Corporation.

Dubowitz, H., Pitts, S. C., Litrownik, A. J., Cox, C. E., Runyan, D., & Black, M. M. (2005). Defining child neglect based on child protective services data. *Child Abuse & Neglect*, *29*, 493–511.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology*, *43*, 1275–1294.

Eisen, M. L., Qin, J., Goodman, G. S., & Davis, S. L. (2002). Memory and suggestibility in maltreated children: Age, stress arousal, dissociation and psychopathology. *Journal of Experimental Child Psychology*, *83*, 167–212.

English, D. J., Upadhyaya, M. P., Litrownik, A. J., Marshall, J. M., Runyan, D. K., Graham, J. C., et al. (2005). Maltreatment's wake: The relationship of maltreatment dimensions to child outcomes. *Child Abuse & Neglect*, *29*, 597–619.

Fivush, R., Bohanek, J., Robertson, R., & Duke, M. (2004). Family narratives and the development of children's emotional well being. In M. W. Pratt & B. H. Fiese (Eds.), *Family stories and the life course: Across time and generations* (pp. 55–76). Mahwah, NJ: Erlbaum.

Gunnar, M. R., & Vazquez, D. M. (2001). Low coritsol and a flattening of expected daytime rhythm: Potential indices of risk in human development. *Development and Psychopathology*, *13*, 515–538.

Gunnar, M. R., & Vazquez, D. (2006). Stress neurobiology and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 533–577). New York: Wiley.

Hart, J., Gunnar, M., & Cicchetti, D. (1996). Altered neuroendocrine activity in maltreated children related to symptoms of depression. *Development and Psychopathology*, *8*, 201–214.

Heffelfinger, A. K., & Newcomer, J. W. (2001). Glucocorticoid effects on memory function over the human lifespan. *Development and Psychopathology*, *13*, 491–513.

Heim, C., Ehlert, U., & Hellhammer, D. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology*, *25*, 1–35.

Heim, C., Newport, J. D., Mletzko, T., Miller, A. H., & Nemeroff, C. B. (2008). The link between childhood trauma and depression: Insights from HPA axis studies in humans. *Psychoneuroendocrinology*, *33*, 693–710.

Heim, C., Owen, M. J., Plotsky, P. M., & Nemeroff, C. B. (1997). Persistent changes in corticotrophin-releasing factor systems due to early life stress: Relationship to the pathophysiology of major depression and posttraumatic stress disorder. *Psychopharmacology Bulletin*, *33*, 185–192.

Heim, C., Plotsky, P. M., & Nemeroff, C. B. (2004). Importance of studying the contributions of early adverse experience to neurobiological findings in depression. *Neuropsychopharmacology*, *29*, 641–648.

Howe, M. L. (1998). Individual differences in factors that modulate storage and retrieval of traumatic memories. *Development and Psychopathology*, *10*, 681–698.

Howe, M. L. (2000). *The fate of early memories: Developmental science and the retention of childhood experiences*. Washington, DC: American Psychological Association.

Howe, M. L., Cicchetti, D., & Toth, S. L. (2006). Children's basic memory processes, stress, and maltreatment. *Development and Psychopathology, 18*, 759–770.

Howe, M. L., Cicchetti, D., Toth, S. L., & Cerrito, B. M. (2004). True and false memories in maltreated children. *Child Development, 75*, 1402–1417.

Howe, M. L., Goodman, G. S., & Cicchetti, D. (Eds.). (2008). *Stress, trauma, and children's memory development: Neurobiological, cognitive, clinical, and legal perspectives*. New York: Oxford University Press.

Howe, M. L., Toth, S. L., & Cicchetti, D. (2006). Memory and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 629–656). New York: Wiley.

Lopez, J. F., Akil, H., & Watson, S. J. (1999). Neural circuits mediating stress. *Biological Psychiatry, 46*, 1461–1471.

Lynch, M., & Cicchetti, D. (1998). Trauma, mental representations, and the organization of memory for mother-referent material. *Development and Psychopathology, 10*, 739–759.

Macfie, J., Cicchetti, D., & Toth, S. L. (2001a). Dissociation in maltreated versus nonmaltreated preschool-aged children. *Child Abuse and Neglect, 25*, 1253–1267.

Macfie, J., Cicchetti, D., & Toth, S. L. (2001b). The development of dissociation in maltreated preschool-aged children. *Development and Psychopathology, 13*, 233–254.

Manly, J. T. (2005). Advances in research definitions of child maltreatment. *Child Abuse & Neglect, 29*, 425–439.

Manly, J. T., Cicchetti, D., & Barnett, D. (1994). The impact of subtype, frequency, chronicity, and severity of child maltreatment on social competence and behavior problems. *Development and Psychopathology, 6*, 121–143.

Manly, J. T., Kim, J. E., Rogosch, F. A., & Cicchetti, D. (2001). Dimensions of child maltreatment and children's adjustment: Contributions of developmental timing and subtype. *Development and Psychopathology, 13*, 759–782.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *Seminars in Medicine of Beth Israel Deaconess Medical Center, 338*, 171–179.

McEwen, B. (2000). Allostasis and allostatic load: Implications for neuropsychopharmacology. *Neuropsychopharmacology, 22*, 108–124.

McEwen, B. S., & Stellar, E. (1993). Stress and the individual mechanisms leading to disease. *Archives of Internal Medicine, 153*, 2093–2101.

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin, 133*, 25–45.

Ogawa, J. R., Sroufe, L. A., Weinfield, N. S., Carlson, E. A., & Egeland, B. (1997). Development of the fragmented self: A longitudinal study of dissociative symptomatology in a normative sample. *Development and Psychopathology, 9*, 855–879.

Pollak, S. D., Cicchetti, D., & Klorman, R. (1998). Stress, memory, and emotion: Developmental considerations from the study of child maltreatment. *Development and Psychopathology, 10*, 811–828.

Pollak, S. D., Cicchetti, D., Klorman, R., & Brumaghim, J. (1997). Cognitive brain event-related potentials and emotion processing in maltreated children. *Child Development, 68*, 773–787.

Pollak, S. D., & Kistler, D. (2002). Early experience alters categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences, USA, 99*, 9072–9076.

Pollak, S. D., Vardi, S., Putzer-Bechner, A. M., & Curtin, J. J. (2005). Physically abused children's regulation of attention in response to hostility. *Child Development, 76*, 968–977.

Porter, C., Lawson, J., & Bigler, E. D. (2005). Neurobehavioral sequelae of child sexual abuse. *Child Neuropsychology, 11*, 203–220.

Putnam, F. W., Helmers, K., & Trickett, P. K. (1993). Development, reliability, and validity of a child disassociation scale. *Child Abuse & Neglect, 17*, 731–741.

Rieder, C., & Cicchetti, D. (1989). Organizational perspective on cognitive control functioning and cognitive-affective balance in maltreated children. *Developmental Psychology, 25*, 382–393.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Sanchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Development and Psychopathology, 13*, 419–450.

Sanchez, M. M., Noble, P. M., Lyon, C. K., Plotsky, P. M., Davis, M., Nemeroff, C. B., et al. (2005). Alterations in diurnal cortisol rhythm and acoustic startle response in nonhuman primates with adverse rearing. *Biological Psychiatry, 57*, 373–381.

Sapolsky, R. M. (1992). *Stress, the aging brain, and the mechanisms of neuron death*. Cambridge, MA: MIT Press.

Sapolsky, R. M. (1996). Stress, glucocorticoids, and damage to the NS: The current state of confusion. *Stress, 1*, 1–19.

Sedlack, A. J., & Broadhurst, D. D. (1996). *The Third National Incidence Study of Child Abuse and Neglect (NIS-3) final report*. Washington, DC: U.S. Department of Health and Human Services.

Shackman, J. E., Shackman, A. J., & Pollak, S. D. (2007). Physical abuse amplifies attention to threat and increases anxiety in children. *Emotion, 7*, 838–852.

Smith, C. A., & Thornberry, T. (1995). The relationship between child maltreatment and adolescent involvement in delinquency. *Criminology, 33*, 451–481.

Susman, E. J., Dockray, S., Schiefelbein, V. L., Herwehe, S., Heaton, J. A., & Dorn, L. D. (2007). Morningness/eveningness, morningto-afternoon cortisol ratio, and antisocial behavior problems during puberty. *Developmental Psychology*, *43*, 811–822.

Tarullo, A. R., & Gunnar, M. R. (2006). Child maltreatment and the developing HPA axis. *Hormones and Behavior*, *50*, 632–639.

Teisl, M., & Cicchetti, D. (2008). Physical abuse, cognitive and emotional processes, and aggressive/disruptive behavior problems. *Social Development*, *17*, 1–23.

Toth, S. L., & Cicchetti, D. (1998). Remembering, forgetting, and the effects of trauma on memory: A developmental psychopathology perspective. *Development and Psychopathology*, *10*, 589–605.

Toth, S. L., Cicchetti, D., Macfie, J., Rogosch, F. A., & Maughan, A. (2000). Narrative representations of moral-affiliative and conflictual themes and behavioral problems in maltreated preschoolers. *Journal of Clinical Child Psychology*, *29*, 307–318.

Toth, S. L., Maughan, A., Manly, J. T., Spagnola, M., & Cicchetti, D. (2002). The relative efficacy of two interventions in altering maltreated preschool children's representational models: Implications for attachment theory. *Development and Psychopathology*, *14*, 877–908.

Valentino, K., Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2008a). Memory, maternal representations, and internalizing symptomatology among abused, neglected, and nonmaltreated children. *Child Development*, *79*, 705–719.

Valentino, K., Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2008b). True and false recall and dissociation among maltreated children: The role of self-schema. *Development and Psychopathology*, *20*, 213–232.

Vazquez, D. M. (1998). Stress and the developing limbic-hypothalamic-pituitary-adrenal axis. *Psychoneuroendocrinology*, *23*, 663–700.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1520–1533

The Contributions of Numerosity and Domain-General Abilities to School Readiness

Lynn S. Fuchs
Vanderbilt University

David C. Geary
University of Missouri

Donald L. Compton, Douglas Fuchs, Carol L. Hamlett, and Joan D. Bryant
Vanderbilt University

Contributions of domain-general and domain-specific numerical competencies were assessed on first graders' number combination skill (NC) and word-problem skill (WP). Students ($n = 205$) between 5 and 7 years of age were assessed on 2 aspects of numerosity, 8 domain-general abilities, NC, and WP. Both aspects of numerosity predicted NC when controlling for domain-general abilities, but domain-general abilities did not account for significant additional variance. By contrast, when controlling for domain-general abilities in predicting WP, only precise representation of small quantities was uniquely predictive, and domain-general measures accounted for significant additional variance; central executive component of working memory and concept formation were uniquely predictive. Results suggest that development of NC and WP depends on different constellations of numerical versus more general cognitive abilities.

In an analysis of six large-scale longitudinal studies, Duncan, Dowsett, Classens, et al. (2008) demonstrated that mathematical competence at school entry predicts mathematics achievement throughout the elementary school years, above and beyond general cognitive ability, classroom attention, social skills, or socioeconomic background. In fact, performance on early mathematical achievement tests was by far the single best predictor of later mathematics achievement. They suggested "it may be beneficial to add domain-specific early skills to the definition of school readiness" (Duncan et al., 2008, p. 1429), but their analysis did not allow for the assessment of which specific mathematical competencies may be the best target for such programs. Candidates center on children's early number sense, including the ability to quickly apprehend the quantities of small sets of items, use counting to determine quantity, estimate the value of large quantities, and intuitively understand the effects of addition and subtraction on quantity (National Mathematics Advisory Panel, 2008).

At the same time, Duncan et al. (2008) showed that classroom attention, a domain-general factor, also predicts later mathematics achievement, above and beyond early mathematical competence. Other studies have revealed that general cognitive ability is also a strong predictor of achievement across academic domains (e.g., Walberg, 1984). General cognitive ability includes working memory capacity, speed of information processing, and logical reasoning (Embretson, 1995; Engle, Tuholski, Laughlin, & Conway, 1999; Kail, 1991), although the relative importance of these domain-general abilities is debated (e.g., Ackerman, Beier, & Boyle, 2005).

In the present study, we focused on the relation between two measures of children's early number sense as well as domain-general cognitive and attentional measures with performance on two critical mathematical abilities: fluency in solving arithmetic problems (e.g., $5 + 7 = 12$; $9 - 5 = 4$), which are sometimes referred to as number combinations (NCs; Baroody, 1985), and simple word problems

---

This research was supported in part by Award RO1 HD053714 and by Core Grant HD15052 from the National Institute of Child Health and Human Development (NICHD) to Vanderbilt University and Award R37 HD045914 cofunded by NICHD and the Office of Special Education and Rehabilitation Services award to the University of Missouri. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Child Health and Human Development or the National Institutes of Health.

Correspondence concerning this article should be addressed to Lynn S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203. Electronic mail may be sent to lynn.fuchs@vanderbilt.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0015

(WPs) that require solution of NCs. By assessing competencies in NC and WP, rather than using broad achievement measures, we explored the relative importance of domain-specific versus domain-general abilities to each of these core aspects of children's early mathematical learning. If Duncan et al. (2008) are correct, then measures of children's number sense should predict NC and WP competence, with domain-general measures accounting for relatively little additional variance. Such findings would help to more firmly establish links between children's informal (before schooling) number sense and their emerging competencies in school-taught mathematics and provide direction for the foci of school readiness programs.

The Domain-General Mechanisms

Of the domain-general mechanisms, working memory, classroom attention, and speed of processing have all been shown to predict NC performance (e.g., Bull & Johnston, 1997; Fuchs et al., 2005; Hitch & McAuley, 1991), but none of the studies assessed these potential mechanisms simultaneously with each other and with domain-specific measures of number sense. Moreover, working memory itself constitutes multiple subcompetencies that may influence NC development in different ways.

At the least, working memory comprises an attention-driven central executive, which controls the representation and manipulation of information in two subsystems that provide short-term storage (Baddeley, 1986). The phonological loop stores auditory or phonological information; the visuospatial sketchpad stores spatial and visual information such as shapes and colors or the location or speed of objects in space. Correlational studies suggest that the central executive is critical for NC development (e.g., Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Swanson & Sachse-Lee, 2001), but the subcomponents of the central executive contributing to this performance, including attentional and inhibitory control, are not well understood. Attentional control may be necessary to maintain the simultaneous activation of problem stems and answers in working memory while counting solutions; inhibitory control may be needed to prevent irrelevant associations from intruding into working memory during problem solving (Barrouillet, Fayol, & Lathuliére, 1997; Engle et al., 1999).

The phonological loop may contribute to the establishment of NCs in long-term memory, because children must encode and maintain accurate phonological representations of problem stems and answers in working memory as they count to solve problems (e.g., Logie & Baddeley, 1987). Evidence on the link between phonological processing and NC skill is, however, inconsistent (e.g., Fuchs et al., 2005; Hecht, Torgesen, Wagner, & Rashotte, 2001; Swanson & Beebe-Frankenberger, 2004). Attentive behavior in the classroom can also influence NC skill (Fuchs et al., 2005; Fuchs et al., 2006), but the relation between this type of attention and the attentional and inhibitory control components of the central executive are not well understood.

The search for the mechanisms underlying NC skill is further complicated by the relation between working memory and speed of processing (the efficiency with which simple tasks are executed; Case, 1985). Bull and Johnston (1997) found that behavioral measurement of processing speed was the best predictor of NC skill among 7-year-olds, subsuming all the variance accounted for by long- and short-term memory, even with reading skill controlled. Hecht et al. (2001) provided corroborating data while controlling for oral vocabulary, and profile analysis (Fuchs, Fuchs, Stuebing, et al., 2008) showed that a behavioral measure of processing speed distinguishes third graders with specific NC deficits from those with WP deficits. However, it is not known whether individual differences in working memory are driven by more fundamental differences in speed of processing (Kail, 1991) or whether the attentional focus associated with the central executive speeds information processing (Engle et al., 1999).

Working memory is also related to WP performance but the contributions of the different components of working memory are not fully understood (e.g., Hitch & McAuley, 1991; Swanson & Sachse-Lee, 2001). Moreover, as noted, performance on working memory measures is correlated with logical reasoning, but processes are distinct to each (Embretson, 1995). General reasoning ability predicts third graders' WP performance even after controlling for working memory (Fuchs et al., 2005, 2006; Swanson & Sachse-Lee, 2001). Finally, we must consider oral language comprehension as a contributor to WP skill, above and beyond the contributions of the phonological loop. Jordan, Levine, and Huttenlocher (1995) demonstrated that kindergarten and first-grade children with oral language impairment perform lower than unimpaired peers on WPs. Fuchs et al. (2006) showed the relation between language and WP skill at third grade while controlling for a variety of other variables, and profile analysis (Fuchs, Fuchs, Stuebing, et al., 2008) revealed that oral language comprehension distinguishes third graders with WP deficits versus those with specific calculation deficits.

The Domain-Specific Mechanisms

In addition to general cognitive abilities, which apply across academic domains, children may have specialized numerical representations and competence in numerical processing (e.g., Baroody, Bajwa, & Eiland, 2009; Feigenson, Dehaene, & Spelke, 2004). Cognitive (Gilmore, McCarthy, & Spelke, 2007), neuropsychological (Domahs & Delazier, 2005), and brain-imaging (Cantlon, Platt, & Brannon, 2009) studies indicate that number sense mechanisms support children's early learning of school-taught mathematics. As per Feigenson et al. (2004), two core systems, one for exact representation of small quantities and one for approximating larger quantities, may be involved in representing numerical value, although controversy exists about whether these systems are supported by distinct cognitive and brain mechanisms (e.g., Cantlon et al., 2009; Gallistel & Gelman, 1992; Mix, Huttenlocher, & Levine, 2002; vanMarle & Wynn, 2009). Whether the underlying mechanisms are cognitively or neurally distinct, they are functionally so; people of any age can rapidly distinguish the quantities of sets of less than four objects but must count or approximate the values of larger sets and Arabic numerals.

The hypothesized first system supports analog representations of approximate numerical magnitudes (Gallistel & Gelman, 1992; Xu, 2003) and can be assessed based on children's placement of numerals on a number line (e.g., Siegler & Booth, 2004). Making placements on a number line, which are based on the use of this approximate magnitude system, results in a pattern that conforms to the natural logarithm (ln) of the numbers (Feigenson et al., 2004; Gallistel & Gelman, 1992; Siegler & Opfer, 2003). In other words, use of this representational system results in placements that are compressed for larger magnitudes such that the perceived distance between 52 and 53 is smaller than the perceived distance between 2 and 3. With schooling, children learn to make placements that conform to the mathematical number line; that is, their placements suggest they understand that the difference between two consecutive numbers is identical regardless of position on the line.

The second hypothesized system enables precise representations of small quantities, the fast apprehension of which is termed *subitizing* (Mandler & Shebo, 1982). With a verbal subitizing task, which involved name identity, Koontz and Berch (1996) asked third and fourth graders with and without a mathematical learning difficulty to determine whether combinations of Arabic numerals (e.g., 3–2), number sets (e.g., ■ ■–■ ■), or numerals and sets were the same (e.g., 2–■ ■) or different (e.g., 3–■ ■). Reaction times for children without mathematics difficulty indicated fast access to representations for quantities of two and three, regardless of whether the code was an Arabic numerical or number set. Children with mathematics difficulty showed fast access to numerosity representations for the quantity of two but appeared to rely on counting to determine quantities of three. Geary et al. (2007; Geary, Bailey, & Hoard, 2009) developed the Number Sets Test to assess related, basic numerical fluency. Children are asked to combine pairs or triplets of Arabic numerals (e.g., 3 4) or sets of objects (e.g., •• ■ ■ ■) and quickly determine if they match a target number (e.g., 5). Verbal subitizing can contribute to fluency if the quantity of small sets of objects is apprehended and then added together (e.g., •• ■ ■ ■ = 5; Geary & Lin, 1998). Children with mathematics difficulty perform poorly on this task compared to peers, independent of general cognitive ability, but performance may be mediated by visuospatial working memory (Geary et al., 2008).

Purpose of the Present Study

In the present study, we assessed the relative contributions of domain-general working memory, attentional, language, and reasoning abilities and domain-specific numerical competencies on individual differences in first graders' NC and WP performance. We had three goals. The first was to determine whether measures of the approximate and exact number representation systems directly affect NC skill when controlling for domain-general abilities. Evidence for one or the other numerosity measure would support Duncan et al.'s (2008) suggestion that domain-specific competence may be important for children's early mathematical learning. Such a finding would also provide much needed information on where to target early interventions for children at risk for poor outcomes in mathematics (e.g., Siegler & Ramani, 2008). Our second goal was to determine how much of the variance in domain-general abilities and numerosity variables represents unique versus shared variance in explaining NC skill. Finding that domain-general abilities uniquely account for much of the variance would suggest that children's early mathematics learning is supported by the same mechanisms that support learning in other domains; finding that the numerosity variables uniquely account for much of the variance would substantiate a domain-specific perspective. By contrast, finding that a substantial amount of the variance is shared would suggest the viability of a combined perspective. Our third goal was to address the same issues as in our first two goals but applied to WPs that require solutions to NCs. These analyses assess whether domain-general or domain-specific relations found for NCs generalize to mathematics content beyond NC skill.

Method

Participants

As part of a prospective 4-year study of the development of NC skill beginning in first grade, data for the present study were collected with the first-year sample at the first assessment wave, when we sampled children from 52 first-grade classrooms in nine Title 1 schools and two non-Title 1 schools (2–9 teachers per school) in a Southeastern metropolitan school district. From 898 students in these classrooms, we secured consent on 580 (64.6%), who were then screened for inclusion using a two-step process. Step 1 involved whole-class screening on the First-Grade Test of Computational Fluency and the First-Grade Test of Mathematics Concepts and Applications (Fuchs, Hamlett, & Fuchs, 1990; see Measures). Teachers first used scripts to administer two practice alternate forms of each measure. Prior to actual screening, teachers returned the unscored practice tests to project staff. Screening occurred with one additional alternate form of each measure. In Step 2, all students who were present on the day of whole-class screening (n = 507) were tested individually on measures of early reading: Rapid Letter Naming (Fuchs et al., 2001) and Word Identification Fluency (Fuchs, Fuchs, & Compton, 2004; see Measures). We indexed early reading for screening because evidence indicates that students with concurrent difficulty with mathematics and reading may experience more severe difficulty with NCs (e.g., Hanich, Jordan, Kaplan, & Dick, 2001).

Using mixture analysis that combined the four screening scores into a latent variable, we created high, average, and at-risk strata. We combined the high and average strata into a single "not-at-risk" group. We then applied the following exclusions. First, we excluded 73 students whose mathematics performance was inconsistent with their overall risk status (i.e., at-risk students with high mathematics scores and not-at-risk students with low mathematics scores). Second, we excluded 121 students whose teachers identified them as not speaking English. Third, we administered the two-subtest Wechsler Abbreviated Intelligence Scale (WASI; Wechsler, 1999) and excluded 7 students with standard scores below 80 on both subtests (i.e., our goal was not to study mental retardation).

From the remaining 306 students, we randomly sampled 240 students, purposely stratifying by the at-risk and not-at-risk strata and ensuring that no more than 8 students came from the same classroom. The present sample comprises the 205 students on whom we had complete data on the study variables. These students had mean standard scores in the normal range on the Reading subtest of the Wide Range Achievement Test–3 (WRAT; Wilkinson, 1993), 105.12 (SD = 15.35); WRAT-Arithmetic, 96.63 (SD = 14.63); KeyMath–Revised (Connolly, 1998)–Numeration, 102.80 (SD = 11.29); KeyMath–Revised–Addition, 99.90 (SD = 10.46); KeyMath–Revised–Subtraction, 97.30 (SD = 11.97); and WASI IQ, 92.70 (SD = 13.31). Of the 205 participants, 109 (53.2%) were boys; 161 (78.5%) received subsidized lunch. Among the students, 112 (54.6%) were African American, 48 (23.4%) were European American, 24 (11.7%) were Hispanic, and 21 (10.2%) were other. Students with a school-identified disability comprised 4.8% of the sample.

Screening Measures to Constitute a Representative Sample

Mathematics screeners. Because these measures were used to screen overall mathematics competence, they included problems with a range of difficulty to be sensitive to the skills of low, average, and high performers.

The First-Grade Test of Computational Fluency (Fuchs et al., 1990) is a one-page test displaying 25 items that sample the range of problems typically found in the first-grade computation curriculum: adding two single-digit numbers (9 items), subtracting two single-digit numbers (10 items), adding three single-digit numbers (2 items), adding two 2-digit numbers without regrouping (2 items), and subtracting a one-digit number from a two-digit number (2 items). Students have 2 min to complete as many items as possible. The score is the number of digits correct (starting from the right, credit is awarded if the correct numeral is in the correct place so that 38 + 26 = 514 would earn one digit correct to reflect incorrect regrouping). Staff entered responses into a computerized program on an item-by-item basis, with 15% of tests reentered by an independent scorer. Data-entry agreement was 99.8% (i.e., number of agreements divided by number of agreement plus disagreement). Coefficient alpha was .97.

The First-Grade Test of Mathematics Concepts and Applications (Fuchs et al., 1990) is a three-page test with 25 items sampling the typical first-grade concepts and applications curriculum (i.e., numeration, concepts, geometry, measurement, applied computation, money, charts, graphs, and WPs). For example, items require students to count arrays of items, write the number after, identify smaller/larger quantities, read clocks, and identify operators within number sentences. The tester reads the words in each item aloud. For 20 items, students have 15 s to respond; for the remaining 5 items, students have 30 s. The score is the number of correct answers. Staff entered responses into a computerized program on an item-by-item basis, with 15% reentered by an independent scorer. Data-entry agreement was 98.8%. Coefficient alpha was .93.

Reading. With Rapid Letter Naming (Fuchs et al., 2001), students are shown a page with 52 letters (26 letters in upper and lowercase) displayed in random order and have 1 min to say letter names. The score is the number of correct letters. If the child finishes before 1 min, the score is prorated. Two-week stability is .94 (Fuchs et al., 2001). With Word Identification Fluency (Fuchs et al., 2004), students have 1 min to read a list of 50 words randomly sampled from a pool of 100 high-frequency preprimer, primer, and first-grade words. If a student completes reading before 1 min, the score is prorated. We administered two alternate forms and averaged the scores. Alternate-form reliability/stability is .97, and correlations with fall and spring Woodcock Reading Mastery Test–Word Identification (Woodcock, 1998) are .77–.82.

Intelligence. Wechsler Abbreviated Intelligence Scale (Wechsler, 1999) measures generalized cognitive ability with Vocabulary and Matrix Reasoning (see description next). Subtest scores are combined to yield an Estimated Full Scale IQ score. Internal consistency reliability exceeds .92.

Domain-General Predictor Measures

Language. We used two tests of language. Woodcock Diagnostic Reading Battery (WDRB)–Listening Comprehension (Woodcock, 1997) measures the ability to understand sentences or passages. With 38 items, students supply the word missing at the end of each sentence or passage. The test begins with simple verbal analogies and associations and progresses to comprehension involving the ability to discern implications. Testing is discontinued after six consecutive errors. The score is the number of correct responses. Internal consistency reliability is .80 at ages 5–18; the correlation with the Woodcock-Johnson Psycho-Educational Battery–Revised (WJ–R) is .73. WASI Vocabulary (Wechsler, 1999) measures expressive vocabulary, verbal knowledge, and foundation of information with 42 items. The first four items present pictures; the student identifies the object in the picture. For the remaining items, the tester says a word that the student defines. Responses are awarded a score of 0, 1, or 2 depending on quality. Testing is discontinued after five consecutive scores of 0. The score is the total number of points. As reported by Zhu (1999), split-half reliability is .86–.87 at ages 6–7; the correlation with the WISC–III Full-Scale IQ is .72.

Concept formation. Woodcock-Johnson Psycho-Educational Battery–Revised (WJ–III) Concept Formation (Woodcock, McGrew, & Mather, 2001) asks students to identify the rules for concepts when shown illustrations of instances and noninstances of the concept. Examinees earn credit by correctly identifying the rule that governs each concept. Cut-off points determine the ceiling. The score is the number of correct responses. As reported by the test developer, internal consistency reliability is .93.

Nonverbal problem solving. Wechsler Abbreviated Intelligence Scale Matrix Reasoning (Wechsler, 1999) measures nonverbal reasoning with four types of tasks: pattern completion, classification, analogy, and serial reasoning. The child looks at a matrix from which a section is missing and completes it by saying the number of or pointing to one of five response options. Examinees earn points by identifying the correct missing piece of the matrix. Testing is discontinued after four errors on five consecutive items or four consecutive errors. The score is the number of correct responses. Internal consistency reliability is .94; the correlation with the WISC–III Full-Scale IQ is .66.

Attentive behavior. The SWAN is an 18-item teacher rating scale (Swanson et al., 2004), sampling items from the *Diagnostic and Statistical Manual of Mental Disorders–IV* (American Psychiatric Association, 1994) criteria for Attention Deficit/Hyperactivity Disorder for inattention (largely distractibility; items 1–9) and hyperactivity/impulsivity (items 10–18). Items are rated as 1 = *far below*, 2 = *below*, 3 = *slightly below*, 4 = *average*, 5 = *slightly above*, 6 = *above*, 7 = *far above*. We report data for the attentive behavior subscale as the average rating across the nine relevant items. We selected this subscale to operationalize attentive behavior, or the ability to maintain focus of attention. The SWAN correlates well with other dimensional assessments of behavior related to attention (http://www.adhd.net). Coefficient alpha in the present study was .97.

Working memory. The Working Memory Test Battery for Children (WMTB–C; Pickering & Gathercole, 2001) comprises nine subtests that assess the central executive, phonological loop, and visuospatial sketchpad. Each subtest has six items at span levels ranging from 1–6 to 1–9. Passing four items at a level moves the child to the next level. At each span level, the number of items to be remembered increases by one. Failing three items terminates the subtest. The order of subtests is designed to avoid overtaxing any one component area of working memory and is generally arranged from easiest to hardest: Digit Recall, Word List Matching, Word List Recall, Nonword List Recall, Block Recall, Mazes Memory, Listening Recall, Counting Recall, and Backward Digit Recall. Each subtest generates a span score and a trials correct score. From these, standard scores are derived for the three component areas.

The *central executive* is assessed with three dual-task items. For Listening Recall, the child determines if a sentence is true or false and then recalls the last word in a series of sentences. For Counting Recall, the child counts a set of 4, 5, 6, or 7 dots on a card and then recalls the number of counted dots at the end of a series of cards. Backward Digit Recall is a standard format backward digit span.

The *phonological loop* comprises Digit Recall, Word List Recall, and Nonword List Recall, which are standard span tasks. The child repeats stimuli spoken by the tester in the same order. In Word List Matching, the tester presents a series of words, beginning with two words and adding one word at each level. The same words, but possibly in a different order, are then presented again, and the child determines if the second list is in the same order as the first list.

The *visuospatial sketchpad* comprises two tasks. Block Recall is another span task, but the stimuli consist of a board with nine raised blocks in what appears as a random arrangement. The blocks have numbers on one side that only the tester can see. The tester taps a block (or series of blocks), and the child duplicates the tapping in the same order. For Mazes Memory, the tester presents a maze with more than one solution and a picture of an identical maze with a path showing one solution. The picture is removed, and the child duplicates the path in the response booklet. At each level, the complexity of the maze increases.

Processing speed. Woodcock-Johnson Psycho-Educational Battery–Revised Visual Matching (Woodcock et al., 2001) asks examinees to locate and circle two identical numbers in rows of six numbers; examinees have 3 min to complete 60 rows and earn credit by correctly circling the matching numbers in each row. As reported by the test developer, internal consistency reliability is .91.

Domain-Specific Measures of Numerosity

We indexed numerosity with Number Line Estimation, which provides an assessment of the approximate magnitude system, and the Number Sets Test, which provides an assessment of precise representation of precise small quantities and the ability to mentally combine these representations (e.g., •• ■■■ = 5).

Approximate numerical magnitudes. Following Siegler and Booth (2004), Number Line Estimation asks students to estimate where numbers fall on a number line. Stimuli are twenty-four 25-cm number lines, displayed across the center of a standard computer screen. Each number line has a start point of 0 and an endpoint of 100 with a target number printed approximately 5 cm above it in a large font (72 pt). Target numbers are 3, 4, 6, 8, 12, 17, 21, 23, 25, 29, 33, 39, 43, 48, 52, 57, 61, 64, 72, 79, 81, 84, 90, and 96. Stimuli are presented in random order for each child. The tester first explains a number line that includes the 0 and 100 endpoints and is marked in increments of 10. After the tester judges that the child recognizes the concept, a number line containing only the endpoints 0 and 100 is presented, and the child points to where 50 should go. A model number line with the endpoints and the location of 50 marked is then presented, and the child compares his or her response to the model. The tester explains how "the number 50 is half of 100, so we put it halfway in between 0 and 100 on the number line." Next, the tester teaches the child to use the arrow keys to place a red pointer on the line where 50 should fall on the computer screen. Then, the measure is administered. For each item, the tester says, "If this is zero (pointing) and this is 100 (pointing), where would you put *N*?" There is no time constraint. We used the absolute value of the difference between the correct placement and the child's placement (i.e., estimation of accuracy), averaged across trials, which the computer automatically calculates. This estimation accuracy score correlates with mathematics achievement (Geary et al., 2007; Siegler & Booth, 2004), and as Siegler and Booth showed, the source of improvement in estimation accuracy is increasing linearity of estimates. Cronbach's alpha on this sample was .91.

Precise representation of small quantities. The Number Sets Test (Geary et al., 2009) assesses the speed and accuracy with which children understand and manipulate small, exact numerosities, while handling those quantities and their Arabic code. Two types of number-set stimuli are used: objects (circles, squares, diamonds, stars) in a 1/2 in. square and an Arabic numeral (18 pt font) in a 1/2 in. square. Stimuli are joined in domino-like rectangles. The two sides of the "domino" are objects–same objects or objects–different objects or Arabic numerals–Arabic numerals or objects–Arabic numerals. These dominos are presented in lines of 5 across a page. The last two lines of the page show 3 three-square dominos. Target sums (5 or 9) are shown in larger font (36 pt) at the top of the page; 5 and 9 were chosen because they represent smaller and larger values within the range of basic Arabic numerals (i.e., 1–9). The first page for each target sum contains objects–same objects and objects–different objects; the second page, objects–Arabic numerals and Arabic numerals–Arabic numerals. On each page, 18 items match the target, 12 are larger than the target, 6 are smaller than the target, and 6 contain 0 or an empty square. The tester begins by explaining two items matching a target sum of 4m then uses the target sum of 3 for practice; with four lines of two dominos, one is a match and one is not. The measure is then administered. The child is told to move across each line of the page from left to right without skipping any; to "circle any groups that can be put together to make the top number, 5 (9)," and to "work as fast as you can without making many mistakes." The child has 60 s per page for the target sum 5, 90 s per page for the target sum 9 (time limits chosen to avoid ceiling effects and to assess fluent recognition of quantities). Signal detection methods are applied to the number of hits and false alarms to generate a *d′* variable that represents the child's sensitivity to quantities (Geary et al., 2009; MacMillan, 2002). Children who correctly identify many target quantities and commit few false alarms have high *d′* scores. Children who have as many hits as false alarms have low *d′* scores, in which case the high number of correct items is due to the child's bias to respond, not sensitivity to quantity.

Mathematics Outcomes

Number combinations. Math Fact Fluency (Fuchs, Hamlett, & Powell, 2003) incorporates two subtests. Addition Fact Fluency comprises 25 addition fact problems with sums from 5 to 12 (2 items have an addend of 1; 1 item has an addend of 0). Subtraction Fact Fluency comprises 25 subtraction fact problems with subtrahends from 5 to 12 (1 item has a minuend of 1; 1 item has a minuend of 0). For each subtest, students have 1 min to write the answers. The score is the number of correct answers across both subtests. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 97.3. Coefficient alpha on this sample was .94.

Word problems. Following Jordan and Hanich (2000), Word Problems comprises 14 brief WPs involving change, combine, compare, and equalize relationships. The NCs embedded in the WPs are sums of 7, 8, or 9 or subtrahends of 6, 7, 8, or 9; there are no addends or minuends of 0 or 1; and answers to the subtraction problems range from 2 to 6. The tester reads each item aloud; students have 30 s to respond and can ask for rereading(s) as needed. The score is the number of correct answers. A second scorer independently rescored 20% of protocols, with agreement of 99.7%. Coefficient alpha on this sample was .84.

Procedure

We assessed students in the fall of first grade. In September, whole-class screening lasted 15 min; individual screening, 5 min. In September and October, we individually tested children on the remaining measures in three 45- to 60-min sessions. Tests were administered by trained examiners, each of whom had demonstrated 100% accuracy in following standard test procedures during mock administrations. All individual sessions were audiotaped, and 15% of tapes, distributed equally across testers, were selected randomly for accuracy checks by an independent scorer. Agreement was between 99.5% and 99.9%. In October, teachers also completed the SWAN Rating Scale on each student.

Results

A principal components factor analysis across WDRB Listening Comprehension and WASI Vocabulary yielded one factor; thus, we created a weighted composite language variable. Each of the remaining constructs was assessed with one measure: for concept formation, WJ–III Concept Formation; for nonverbal problem solving, WASI Matrix Reasoning; for attentive behavior, SWAN; for working memory, the phonological loop, and the visuospatial sketchpad, summary scores from the WMTB–C subtests; for processing speed, WJ–III Visual Matching. The numerosity measures were Number Line Estimation and the Number Sets Test. The two mathematics outcomes were NCs and WPs. In Table 1, we show raw and standard score means and standard deviations, along with zero-order correlations.

Predictors of Number Combination Skill

We contrasted two models in predicting individual differences in NC skill (see Table 2). In Model 1-NC, we tested the effects of the two numerosity measures, entering Number Line Estimation and the Number Sets Test simultaneously into the regression analysis. Each was a significant predictor in the presence of the other. Overall, Number Line Estimation (negative scores represent accurate placements) and the Number Sets Test (i.e., d' score) explained 65.6% of the variance in NC skill, $F(2, 202) = 192.75$, $p < .001$. Partitioning this variance revealed that 36.9% of the explained variance was shared between Number Line Estimation and the Number Sets Test, 58.5% was unique to Number Sets Test, and 4.6% was unique to Number Line Estimation.

Model 2-NC shows that each numerosity measure (Number Line Estimation and the Number Sets Test) was a significant predictor of NC skill in the presence of the eight domain-general abilities. The only domain-general ability that was a uniquely significant predictor of NC skill was attentive behavior. Overall, the two numerosity measures and the eight domain-general abilities explained 67.7% of the variance in NC skill; the addition of the eight domain-general abilities did not significantly increase R^2: 2.1% increase, F change $(8, 194) = 1.56$, $p = .141$. Partitioning Model 2-NC variance between the numerosity measures and the domain-general abilities revealed that 62.0% of the

Table 1
Means, Standard Deviations, and Correlations Among Domain-General Ability, Numerosity, and Mathematics Outcome Variables (n = 205)

	Raw score	Standard score	Language		Correlations													
Variables	*M*	*SD*	*M*	*SD*	LF	LC	V	CF	NP	IA	EF	PL	VS	PS	NS	NL	NCs	WP
Domain-general abilities																		
Language factor (LF)	0.01	1.01	—	—	—													
WDRB listening comprehension (LC)	14.00	5.16	87.18	16.34	88	—												
WASI vocabulary (V)	17.25	6.65	41.22	10.50	87	61	—											
Concept formation (CF)	9.42	5.47	95.00	13.44	52	54	44	—										
Nonverbal problem solving (NP)	8.70	5.43	49.01	9.57	42	36	42	50	—									
Attentive behavior (IA)	38.99	12.65	—	—	49	46	45	47	42	—								
WMTB executive function (EF)	—	—	76.24	18.28	55	50	42	48	51	48	—							
WMTB phonological loop (PL)	—	—	96.60	17.65	43	39	25	33	27	31	49	—						
WMTB visuospatial sketchpad (VS)	—	—	78.56	15.14	30	24	21	31	42	32	47	32	—					
Processing speed (PS)	474.37	7.28	95.80	15.38	36	36	35	43	43	44	44	24	37	—				
Numerosity measures																		
Number sets (NS)	0.05	1.21	—	—	47	45	47	59	51	59	53	28	35	51	—			
Number line estimation (NL)	22.48	7.75	—	—	45	37	51	34	40	39	37	14	27	36	47	—		
Mathematics outcomes																		
Number combinations (NCs)	20.44	12.15	—	—	48	43	49	54	46	58	47	30	35	50	79	52	—	
Word problems (WP)	3.83	3.04	—	—	54	49	51	58	49	52	59	31	34	42	65	45	66	

Note. Standard scores are $M = 100$ ($SD = 15$) except for Wechsler Abbreviated Intelligence Scale Vocabulary and nonverbal problem solving, where $M = 50$ ($SD = 10$). Multiply correlations by .01. Correlations involving Number Line Estimation are negative. WASI = Wechsler Abbreviated Intelligence Scale; WDRB = Woodcock Diagnostic Reading Battery; WJ–III = Woodcock-Johnson Psycho-Educational Battery–Revised; Language = WDRB Listening Comprehension and WASI Vocabulary; Concept Formation = WJ–III Concept Formation; Nonverbal problem solving = WASI Matrix Reasoning; WMTB = Working Memory Test Battery–Children; Attentive behavior = the SWAN; Processing speed = WJ–III Visual Matching; Arithmetic = the Test of Mathematics Fact Fluency.

Table 2
Regression Models Predicting Individual Differences in First-Grade Number Combination Skill

Model	R^2	*SEE*	R^2 change	*F* change	*p* value	*B*	*SE*	β	*t*	*p* value
1. Numerosity	.656	7.16	.656	(2, 202) = 192.75	< .001					
Constant						27.03	1.73		15.66	< .001
Number line estimation						−0.31	0.07	−.20	−4.21	< .001
Number sets test						7.06	0.47	.70	15.02	< .001
2. Numerosity + domain general	.677	7.08	.021	(8, 194) = 1.56	.141					
Constant						14.64	6.03		2.43	.016
Language						0.43	0.69	.04	0.62	.535
Attentive behavior						0.11	0.05	.12	2.18	.031
Central executive						−0.41	0.04	−.06	−1.03	.305
Phonological loop						0.04	0.03	.05	1.06	.290
Visuospatial sketchpad						0.03	0.04	.04	0.74	.459
Concept formation						0.03	0.05	.03	0.53	.594
Nonverbal reasoning						−0.03	0.07	−.02	−0.37	.712
Processing speed						0.03	0.04	.04	0.85	.394
Number line estimation						−0.27	0.08	−.17	−3.46	.001
Number sets test						6.11	0.59	.61	10.37	< .001

Table 3
Regression Models Predicting Individual Differences in First-Grade Word-Problem Skill

Model	R^2	*SEE*	R^2 change	*F* change	*p* value	*B*	*SE*	β	*t*	*p* value
1. Numerosity	.444	2.28	.444	(2, 202) = 80.66	< .001					
Constant						5.47	0.54		10.20	< .001
Number line estimation						−0.76	0.23	−.20	−3.37	.001
Number sets test						1.39	0.15	.55	9.49	< .001
2. Numerosity + domain general	.552	2.09	.108	(8, 194) = 5.82	< .001					
Constant						−1.04	1.78		−0.58	.560
Language						0.37	0.20	.12	1.83	.069
Attentive behavior						0.02	0.02	.08	1.26	.209
Central executive						0.04	0.01	.23	3.30	.001
Phonological loop						−0.01	0.01	−.03	−0.56	.578
Visuospatial sketchpad						−0.00	0.01	−.01	−0.09	.929
Concept formation						0.03	0.01	.14	2.24	.026
Nonverbal reasoning						0.01	0.02	.02	0.34	.736
Processing speed						−0.01	0.01	−.04	−0.73	.467
Number line estimation						−0.04	0.02	−.10	−1.66	.100
Number sets test						0.80	0.17	.32	4.63	< .001

explained variance was shared, 34.9% was unique to the numerosity measures, and 3.1% was unique to domain-general abilities.

Predictors of WP Skill

We followed the same procedure in predicting individual differences in WP skill (see Table 3). In Model 1-WP, each numerosity measure (Number Line Estimation and the Number Sets Test) was a significant predictor in the presence of the other. Together, Number Line Estimation and the Number Sets Test explained 44.4% of the variance in WP skill, $F(2, 202) = 80.66$, $p < .001$. Partitioning this variance revealed that 38.7% of the explained variance was shared between Number Line Estimation and the Number Sets Test, 55.2% was unique to d', and 6.1% was unique to Number Line Estimation.

In Model 2-WP, the Number Sets Test but not Number Line Estimation was a significant predictor of WP skill in the presence of the eight domain-general abilities. The domain-general abilities that were uniquely significant predictors of WP skill were the central executive component of working memory and concept formation, with a trend for language ability ($p = .069$). Overall, the numerosity measures and the eight domain-general abilities explained 55.2% of the variance in WP skill, with the addition of the domain-general abilities resulting in a substantial (10.8%) and significant increase in R^2, F change(8, 194) = 5.82, $p < .001$. Partitioning Model 2-WP variance between the numerosity measures and the domain-general abilities revealed that 68.3% of the explained variance was shared, 12.1% was unique to the numerosity measures, and 19.6% was unique to domain-general abilities.

Discussion

We assessed the relative contributions of domain-general abilities and domain-specific numerical competencies on individual differences in first graders' NC and WP skill. We focused first on NC skill because the National Mathematics Advisory Panel (2008) and the National Council of Teachers of Mathematics (2006) identified NC fluency as a core arithmetic skill, and because attaining NC fluency is difficult for many children (e.g., Goldman, Pellegrino, & Mertz, 1988).

We began by comparing the relative contributions of measures representing the two hypothesized number sense systems to the prediction of individual differences in NC skill. We entered a measure of approximate representation of large quantities (i.e., Number Line Estimation; Siegler & Booth, 2004) and a measure of exact representation of small quantities (i.e., d' from the Number Sets Test; Geary et al., 2009) simultaneously into regression analysis predicting NC skill. Each was a significant predictor in the presence of the other, and the two numerosity measures together accounted for a large proportion of the variance (65.6%). This suggests that both core systems hypothesized to comprise number sense contribute to the development of NC skill. Even so, most (58.5% of the 65.6%) of this variance was unique to the measure of exact representations of small quantities (i.e., the Number Sets Test), with 36.9% of the explained variance shared between the two numerosity constructs and only 4.6% attributable specifically to approximate representation of large quantities (i.e., Number Line Estimation).

The importance of the approximate system of number in NC skill development is in keeping with findings of some recent studies (Gilmore et al., 2007; Halberda, Mazzocco, & Feigenson, 2008). Yet, our results suggest that fluency in accessing exact representations of small quantities may be more essential. This makes sense given that NC skill and items on the Number Sets Test require quick recognition and combination of small magnitudes. One possibility is that children who have a strong foundation for verbally subitizing and adding the corresponding quantities are at an advantage in learning NC. This might explain Koontz and Berch's (1996) finding that children with mathematics disability have a smaller subitizing range; it may also explain the common finding that children with mathematics disability have difficulty learning NCs (Geary, 1993). At the same time, it is likely that for other types of mathematics performance, approximate representation of large quantities is the more essential system. In that vein, Booth and Siegler (2008) provided data suggesting the potential importance of Number Line Estimation in predicting students' learning of an addition estimation task. The authors did not, however, consider exact representation of small quantities as a competing predictor. In future studies, researchers might simultaneously consider the two hypothesized core systems of number sense when investigating various aspects of mathematics performance, as done in the present study. In any case, our results support Duncan et al.'s (2008) conclusion about the importance of early numerical competence for supporting formal mathematical learning in school.

After examining the effects of the two numerosity measures on NC skill, we contrasted a model that assessed the effects of the numerosity measures in the presence of the eight domain-general abilities. In this much more stringent test, both numerosity measures retained their status as unique predictors. On this basis, we conclude that each of the two hypothesized core systems of number sense directly affects the development of NC skill even when controlling for the variance associated with a host of domain-general abilities. Just as interesting, however, results showed that the addition of the domain-general abilities resulted in only a small, nonsignificant (2.1%) increase in R^2. This was the case even though several of our measures of domain-general abilities (i.e., central executive, phonological loop, and processing speed) involved numerals. As per Landerl, Bevan, and Butterworth (2004), this creates a stringent test of the domain-specific hypothesis due to the potential for confounding domain-general abilities with numerical processing abilities. In a similar way, because we excluded 73 students due to mathematics performance that was inconsistent with their overall risk status, it is possible that results are biased toward a domain-general perspective, creating an even more stringent test of the domain-specific hypothesis.

Yet, our findings lend support for a domain-specific perspective on NC skill, not only indicating that both hypothesized aspects of numerosity are important in NC skill but also that (a) NC skill does not depend on domain-general abilities above and beyond the influence of numerosity and (b) domain-general abilities do not mediate the effects of numerosity on NC skill. This suggests that NC skill grows out of the development of meaningful and interconnected knowledge about numbers that may have their foundation in children's early ability to subitize and estimate magnitude (e.g., Baroody et al., 2009). In the present study, we did not address whether individual differences in numerosity result from innate differences or from the nature of socially mediated preschool experiences. Regardless of how these individual differences develop, results indicate the importance of providing strong preschool opportunities to help children construct a strong sense of numerosity.

At the same time, however, we also examined how much of the variance in domain-general abilities and numerosity variables is unique versus shared when explaining NC skill. Partitioning the variance revealed that although 34.9% of the explained variance was unique to the numerosity measures (and only 3.1% was unique to domain-general abilities), 62.0% of the explained variance was shared. Finding that a substantial amount of variance is shared indicates the need for additional research on the viability of a combined perspective. Geary et al. (2009) demonstrated that although performance on the Number Sets Test appears to capture fluency in processing basic numerical information, it is influenced by the central executive and general cognitive ability. Whether this represents method variance in the construction of the Number Sets Test or whether working memory and cognitive ability are important for the functional use of domain-specific number representations during mathematics learning remains to be determined.

Our last question centered on whether the relations for NC skill apply to WP performance, which requires but is not limited to NC skill. When Number Line Estimation and the Number Sets Test were simultaneously entered into regression analysis to predict WP performance, each was a significant predictor in the presence of the other, as was the case for NC skill. However, these numerosity variables accounted for less (44.4%) of the explained variance in WP skill compared to NC performance (65.6%). So it was not surprising that when we assessed the effects of the numerosity measures in the presence of the eight domain-general abilities, this more stringent test revealed that only one of these core systems of number sense, exact representation of small quantities (i.e., *d′* of the Number Sets Test), not approximate representation of large quantities (i.e., Number Line Estimation), was a unique predictor. This indicates that only exact representation of small quantities exerts a direct effect on WP skill whereas the effects of approximate representation of large quantities are mediated by the domain-general abilities.

In fact, the addition of the set of domain-general measures resulted in a substantial (10.8%) and significant increase in explained variance in WP skill. This stands in stark contrast to the finding that domain-general abilities together did not improve the prediction of NC performance beyond what was captured by the numerosity measures. Moreover, partitioning the variance between the numerosity measures and the domain-general abilities revealed that more variance was attributable to domain-general abilities (19.6%) than to the numerosity measures (12.1%). By contrast, in predicting NC skill, the corresponding percentages were 3.1 and 34.0. In line with the prediction of NC skill, however, a hefty portion (68.3%) of the explained variance was shared between numerosity and domain-general abilities, again suggesting additional research on the viability of a combined perspective.

In conducting research to understand which domain-general abilities may underlie WP skill, our findings indicate the importance of three abilities. First, the central executive component of working memory was uniquely predictive of WP performance, even when two numerosity and seven other domain-general abilities were competing for variance. Theoretical frameworks posit that WPs involve construction of a problem model that requires working memory (e.g., Kintsch & Greeno, 1985). Much of the literature, including the present results, provides support for this theoretical framework (e.g., Fuchs et al., 2005; Swanson & Beebe-Frankenberger, 2004; Swanson & Sachse-Lee, 2001), although findings are not consistent (e.g., Fuchs et al., 2006; Fuchs, Fuchs, Stuebing, et al., 2008; Vukovic & Siegal, 2010).

A second potentially important domain-general ability in determining WP skill, as reflected in the present study, is concept formation. As with the central executive, concept formation was uniquely predictive of WP performance, even when two numerosity and seven other domain-general abilities were competing for variance. Concept formation involves identifying, categorizing, and determining rules to classify objects and therefore appears to represent general reasoning or problem-solving ability, as is required in WPs. In prior research, concept formation has also emerged as a determinant of WP skill (Fuchs et al., 2006; Fuchs, Fuchs, Stuebing, et al., 2008; Swanson & Sachse-Lee, 2001). Additional studies are needed to determine whether concept formation predicts ease of learning and transferring WP schemas (e.g., Fuchs, Fuchs, Craddock, et al., 2008; Sweller, 2004).

The present results also suggest the potential importance of a third domain-general ability: language. Although it only approached statistical significance ($p = .069$), we nonetheless discuss this finding because prior studies strongly support the role of language in WP performance, thereby suggesting the appropriateness of a one-tailed test of significance. Jordan et al. (1995) documented the importance of language when kindergarten and first-grade language-impaired children performed lower than unimpaired peers on WPs. Fuchs et al. (2006) showed its relation to WP skill among third graders while controlling for a host of other variables, and profile analysis (Fuchs, Fuchs, Stuebing, et al., 2008) demonstrated that poor language ability was the domain-general ability that reliably distinguished third graders with WP deficits from those with specific calculation deficits. Moreover, it makes sense that language supports WP performance given the obvious need to process linguistic information when building representations of WPs.

In a more general way, the pattern of findings in the present study, whereby the role of numerosity and domain-general abilities differed across NC and WP skill, also suggests that not all aspects of mathematics performance tap the same underlying abilities. These findings lend support for the notion that NC and WP skill may represent distinct domains of mathematical cognition. Results echo Fuchs, Fuchs, Stuebing, et al. (2008), in which third-grade students at the lower end of the performance range in calculations versus WPs showed distinctive cognitive profiles, and Hart, Petrill, Thompson, and Plomin (2009), who found that mathematics problem solving has different genetic and environmental influences than mathematics calculation. Together, these studies suggest that the development of mathematical cognition does not depend on a uniform set of numerical or general cognitive competencies but rather that the relative contributions of numerosity and domain-general explanations to the development of mathematical cognition differ depending on the nature of mathematics performance. In fact, elementary-school mathematics is conceptualized in strands (e.g., concepts, numeration, measurement, arithmetic, algorithmic computation, and problem solving), and high school curriculum offerings include algebra, geometry, trigonometry, and calculus. Yet, little is understood about how aspects of mathematical cognition relate to each other: which are shared or distinct or how development in one domain corresponds with development in another or how learning in one domain promotes learning in another. Research is needed to expand theoretical insight into the nature of these types of mathematical competence and their development.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baroody, A. J. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal of Research on Mathematics Education*, *16*, 83–98.

Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews*, *15*, 69–79.

Barrouillet, P., Fayol, M., & Lathuliére, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning disabilities. *International Journal of Behavioral Development*, *21*, 253–275.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*, 1016–1031.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, *65*, 1–24.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Science*, *13*, 83–91.

Case, R. (1985). *Intellectual development: Birth to adulthood*. San Diego, CA: Academic Press.

Connolly, A. J. (1998). *KeyMath–Revised*. Circle Pines, MN: American Guidance Service.

Domahs, F., & Delazier, M. (2005). Some assumptions and facts about arithmetic facts. *Psychology Science*, *47*, 96–111.

Duncan, G. J., Dowsett, C. J., Classens, A., Magnusson, K., Huston, A. C., Klebanov, P., et al. (2008). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence, 20,* 169–189.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *TRENDS in Cognitive Science, 8,* 307–314.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71,* 7–21.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29–43.

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology, 100,* 491–509.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. E. (2008). Problem-solving and computation skills: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100,* 30–47.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., et al. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93,* 251–267.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *First grade test of computational fluency and first-grade test of mathematics concepts and applications.* Nashville, TN: Vanderbilt University.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Math fact fluency.* Nashville, TN: Vanderbilt University.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362.

Geary, D. C. (2007). An evolutionary perspective on learning disability in mathematics. *Developmental Neuropsychology, 32,* 471–519.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment, 27,* 265–279.

Geary, D. C., Hoard, M. K., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1459.

Geary, D. C., & Lin, J. (1998). Numerical cognition: Age-related differences in the speed of executing biologically-primary and biologically-secondary processes. *Experimental Aging Research, 24,* 101–137.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447,* 589–591.

Goldman, S. R., Pellegrino, J. W., & Mertz, D. L. (1988). Extended practice of addition facts: Strategy changes in learning-disabled students. *Cognition and Instruction, 5,* 223–265.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455,* 665–668.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology, 93,* 615–626.

Hart, S. A., Petrill, A. S., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101,* 388–402.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computational skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79,* 192–227.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning disabilities. *British Journal of Psychology, 82,* 375–386.

Jordan, N. C., & Hanich, L. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33,* 567–578.

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *Journal of Learning Disabilities, 28,* 53–64.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490–501.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109–129.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, 2, 1–23.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9 year old students. *Cognition*, *93*, 99–125.

Logie, R. H., & Baddeley, A. D. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 310–326.

MacMillan, N. A. (2002). Signal detection theory. In J. Wixted & H. Pashler (Eds.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (Vol. 4, 3rd ed., pp. 43–90). New York: Wiley.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology General*, *111*, 1–21.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Math without words: Quantitative development in infancy and early childhood*. New York: Oxford University Press.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Reston, VA: National Council of Teachers of Mathematics. Retrieved July 6, 2010, from http://www.nctm.org/Catalog/product.aspx?id=13089

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC: United States Department of Education. Retrieved from http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children*. London: Psychological Corporation.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444.

Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Siegler, R. S., & Ramani, G. B. (2008). Playing board games promotes low-income children's numerical development. *Developmental Science*, *11*, 655–661.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious mathematics difficulties. *Journal of Educational Psychology*, *96*, 471–491.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*, 294–321.

Swanson, J., et al. (2004). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales. Retrieved December 20, 2004, from http://www.adhd.net

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*, 9–31.

vanMarle, K., & Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. *Cognition*, *111*, 302–316.

Vukovic, R. K., & Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth grade. *Learning Disabilities Research and Practice*, *25*, 25–38.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, *41*, 19–27.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3*. Wilmington: Wide Range.

Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Test–Revised*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representation. *Cognition*, *89*, B15–B25.

Zhu, J. (1999). *WASI manual*. San Antonio, TX: Psychological Corporation.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1534–1549

Children's Classroom Engagement and School Readiness Gains in Prekindergarten

Nina C. Chien and Carollee Howes
University of California at Los Angeles

Margaret Burchinal
University of California at Irvine

Robert C. Pianta
University of Virginia

Sharon Ritchie, Donna M. Bryant,
Richard M. Clifford, Diane M. Early
and Oscar A. Barbarin
University of North Carolina at Chapel Hill

Child engagement in prekindergarten classrooms was examined using 2,751 children (mean age = 4.62) enrolled in public prekindergarten programs that were part of the Multi-State Study of Pre-Kindergarten and the State-Wide Early Education Programs Study. Latent class analysis was used to classify children into 4 profiles of classroom engagement: *free play*, *individual instruction*, *group instruction*, and *scaffolded learning*. *Free play* children exhibited smaller gains across the prekindergarten year on indicators of language/literacy and mathematics compared to other children. *Individual instruction* children made greater gains than other children on the Woodcock Johnson Applied Problems. Poor children in the *individual instruction* profile fared better than nonpoor children in that profile; in all other snapshot profiles, poor children fared worse than nonpoor children.

There is growing concern about children's lack of readiness for school (Bowman, Donovan, & Burns, 2000). Evidence suggests that children's school readiness, particularly for children from disadvantaged backgrounds, is enhanced in prekindergarten programs during the year before kindergarten (e.g., Magnuson, Meyers, Ruhm, & Waldfogel, 2004). About three fourths of the states now offer such programs (Barnett, Hustedt, Friedman, Boyd, & Ainsworth, 2007). Identification of program characteristics related to improving skills is important given the huge investment of tax dollars and the importance of having vulnerable children enter school ready to learn. The current study describes patterns of children's engagement in prekindergarten classrooms. Specifically, we used person-centered analyses to group children into profiles (i.e., groups) of classroom engagement that reflected the dominant literatures on early childhood education. Then, we explored whether profile membership was linked to gains in school readiness during the prekindergarten year. Additionally, we examined whether some profiles were particularly beneficial for poor children.

Classroom Quality: Children's Classroom Engagement Instead of Classroom Environment

Higher quality prekindergarten programs are associated with more positive child outcomes (e.g., Burchinal et al., 2000). The literature on child-care environmental quality is primarily based on both constructivist theory, in which the adults' role is to provide children with rich materials that promote child-initiated exploration (Ginsburg & Opper, 1988), and sociocultural theory, in which the adults'

The NCEDL Multi-State and SWEEP Study of Pre-Kindergarten were conducted by a team of researchers, including Oscar A. Barbarin, Donna M. Bryant, Margaret Burchinal, Richard M. Clifford, Diane M. Early, Carollee Howes, and Robert C. Pianta. This study is supported under the Educational Research and Development Center Program, PR/Award R307A60004, as administered by the Institute of Education Sciences, U.S. Department of Education. However, the contents do not necessarily represent the positions or policies of the U.S. Department of Education, and endorsement by the federal government should not be assumed. NCEDL is grateful for the help of the many children, parents, teachers, administrators, and field staff who part of this study.

Correspondence concerning this article should be addressed to Nina C. Chien, University of California, San Diego, 9500 Gilman Drive, #0927, La Jolla, CA 92093-0927. Electronic mail may be sent to nchien@ucsd.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0016

role is to provide frequent sensitive and responsive interactions with children (Vygotsky, 1962). The assessment of quality is typically at the classroom level and involves global assessments across all activities. Many measures focus on the teacher, although some widely used measures examine the relationship between an individual child and teacher.

Examining children's classroom engagement can provide additional information that is not available in the environmental quality assessments. The importance of focusing on child-level experiences is stressed in Howes's theoretical model for children's child-care experiences (Howes, 2000), where children's experiences—play activities, peer play, and relationships—are embedded within the context of the classroom. That is, although measures of child-care environmental quality describe the context within which child engagement occurs, a direct examination of child engagement at the child level is needed to understand exactly what children are doing—what activities are occupying children's time—within those contexts. Although children's engagement and the classroom social context certainly influence each other (e.g., Howes & Smith, 1995), they remain distinct constructs.

Another reason to look beyond environmental quality to classroom engagement is that common measures of classroom environmental quality might not adequately capture dimensions of the classroom that are linked to child outcomes. A meta-analysis of 20 studies of the relation between measures of child-care quality and child outcomes showed that the relation is weak (r = .12; Burchinal et al., 2008). The types of activities and level of interactions with the caregiver are also known to be important in predicting preschool children's learning in child-care settings (Howes et al., 2008).

A Person-Centered Approach

Child engagement has thus far been examined using regression, which examines each child engagement activity individually (e.g., Howes et al., 2008). Another method is a person-centered approach, which would consider the entire spectrum of activity engagement of each child to place children into profiles, or subgroups. As such, children within the same profile would exhibit more similar patterns of classroom engagement than children in different profiles. A person-centered approach is more holistic than regression because it considers the entire constellation of child engagement rather than one child engagement at a time.

One type of person-centered analysis, cluster analysis, was used in an earlier study that clustered children based on their classroom engagement (Tonyan & Howes, 2003). Latent class analysis (LCA) is another person-centered approach that offers several advantages over cluster analysis. One advantage is that LCA provides model fit statistics that allow assessment of the model fit to the data, and the appropriateness of the number of profiles specified. In addition, because LCA is model based, the same results can theoretically be replicated with an independent sample (Muthén & Muthén, 2000). LCA would be an ideal method for identifying profiles of children based on their patterns of classroom engagement.

Patterns of Classroom Engagement

Based on dominant models of early childhood education (e.g., the constructivist or sociocultural models), three profiles might be expected to emerge from a LCA of children's classroom engagement: a profile that emphasizes free-choice play and exploration, a profile that emphasizes teacher instruction, and a profile that emphasizes teacher scaffolding.

One model of early childhood education encourages child-directed exploration in activities that the child chooses, positing that many developmental competencies are acquired only through play (Johnson, Christie, Yawkey, & Wardle, 1987). For example, sociodramatic play develops language and problem-solving skills; constructive play (e.g., blocks) helps children learn about symmetry and practice making mental plans. These skills, according to this model, are not readily learned via teacher instruction. Based on this model, we expect to see a profile of children that engages in primarily free-choice play activities.

A second model of early childhood education posits that children learn most from teacher instructional support, defined as large amounts of literacy instruction, high-quality teacher feedback, and teacher-led discussions that elicit cognitive skills (Hamre & Pianta, 2005). One study reported that more class time spent on direct and explicit instruction involving teacher feedback was linked to higher levels of student achievement (Meyer, Wardrop, Hastings, & Linn, 1993; Pianta, La Paro, Payne, Cox, & Bradley, 2002). Another study found that increased instructional time in each of four preacademic activities—letter-sound, oral language, being read to, and mathematics—was associated with higher teacher ratings of children's language and literacy skills (Howes et al., 2008). Based on this model, we expect to see a profile of children that spends a lot of time in teacher-directed instructional activities.

A third model of learning emphasizes the importance of "scaffolding" by a more knowledgeable other that enables children to think and complete tasks at a higher level than if they were unassisted (Wood, Bruner, & Ross, 1976). An important element of quality early childhood education is teacher scaffolding of children's learning (Smith, 1996), including scaffolding to develop young children's literacy (Henderson, Many, Wellborn, & Ward, 2002). Because whether scaffolding occurs during free play or teacher instruction is less important, we might expect a third profile of children who are not distinguishable from other children by their engagement in free play or instructional time, but rather in their receipt of teacher scaffolding.

Different models of classroom engagement may be more beneficial for different types of child outcomes. Children in classrooms that give children opportunities for free play and exploration might have more developed language and advanced mathematics and spatial skills (Johnson et al., 1987), according to the literature on free-choice play. Children in classrooms utilizing the instructional model might be expected to make greater gains in basic academic skills such as knowing numbers and letters and how to write their names, because these are skills more readily taught via teacher instruction and less likely learned through free play. Children in classrooms utilizing scaffolding might have more developed literacy skills (Henderson et al., 2002) or problem-solving skills (Rogoff, 1990). Including a variety of child outcomes makes it possible to assess whether certain models of early childhood education are best suited for particular domains of learning.

Child Engagement and Poverty Status

Children's classroom engagement may also be linked to their poverty status. One study found, via teacher interviews, notable differences in the reported beliefs of preschool teachers of lower versus middle socioeconomic status (SES) children (Lee & Ginsburg, 2007). Teachers of lower SES children reported explicitly focusing on developing children's literacy and mathematics skills through direct instruction to prepare children for kindergarten. Teachers of higher SES children, on the other hand, reported developing skills more indirectly by allowing children to engage in the literacy or mathematics activities of their choice in a classroom environment filled with rich learning materials. We therefore expect children's classroom engagement to differ by poverty status, consistent with teacher beliefs reported in this study.

If preschool teachers of low- versus middle-income children have different beliefs about what constitutes an ideal preschool education, is there any evidence that their beliefs are accurate? That is, do low-income children indeed benefit more from a focus on mathematics and literacy skill development, whereas middle-income children benefit more from the freedom to choose activities? One study found that children who were at risk of school failure benefited more from classrooms with more direct literacy instruction, evaluative feedback, and instructional conversations during teacher-led discussions, compared to children who were not at risk (Hamre & Pianta, 2005). Therefore, poor children might be expected to benefit more from direct instruction of academic skills than from child-initiated free-play activities.

In the current study, we used LCA to classify children into profiles based on their classroom engagement. We then used profile membership to predict children's gains from fall to spring of the prekindergarten year, in the areas of language and literacy and mathematics. Next, we examined sociodemographic differences across the class engagement profiles. Finally, we assessed whether different child engagement profiles provided different amounts of gains for poor versus nonpoor children.

Method

Participants

Study data come from the National Center for Early Development and Learning Multi-State Study of Pre-Kindergarten, and a follow-up study, the State-Wide Early Education Programs Study (SWEEP). Both studies used a stratified random sampling method to select programs within states, classrooms within programs, and children within classrooms. States with large numbers of children enrolled in public prekindergarten and with programs that have had time to mature were selected. The Multi-State included six states (California, Illinois, Georgia, Kentucky, New York, and Ohio), with 40 programs selected from each state, and began in fall of 2001; the SWEEP included five states (Massachusetts, New Jersey, Texas, Washington, and Wisconsin), with 100 programs selected from each state, and began in fall of 2003. Programs were diverse with regard to urbanicity. All programs were state funded, although funding streams differed within and across states; 15% of programs were part of Head Start programs. A total of 701 programs participated, and one classroom was randomly selected from each program. In both the Multi-State and the SWEEP, 94% of classroom teachers agreed to participate. In the Multi-State and SWEEP, 61% and 55% of parents, respectively, gave consent to participate. Of the children with parent consent, four children (two boys and two girls) from each classroom were randomly selected to participate, making for a total of 2,966 children.

Sample classrooms had an average class size of 19 children and an adult–child ratio of 1:8.6. A majority of the children (58%) were from families living below the federal poverty line, and about half were boys (49%). The sample included children who were European American (41%), African American (18%), Latino (27%), Asian American (4%), Native American (1%), and of other ethnicities (10%). Mean maternal education was 12.8 years. Children with no classroom engagement observations ($n = 215$) were dropped from the study, resulting in a final sample of 2,751 children.

Measures

Classroom Observations

The Emerging Academics Snapshot (Ritchie, Howes, Kraft-Sayre, & Weister, 2001) is a measure of children's classroom engagement that captures children's moment-to-moment activities. Observations were conducted over 2 days in the Multi-State and over 1 day in the SWEEP, each in the spring. Each child was observed in 20-s interval "snapshots," followed by a 40-s coding period. The data collector then observed each of the other study children in that classroom before coming back to observe the first child again, continuing in this manner for the entire day. During each 20-s snapshot, children were coded with one of six mutually exclusive *activity settings*: basics, free choice, individual time, meals, small group, and whole group; one or more *preacademic engagements*: esthetics, fine motor skills, gross motor skills, letter and sound, mathematics, oral language development, prereading, read to, science, social studies, and writing; and one or more *teacher–child interactions*: routine, minimal, simple, elaborated, scaffolding, and didactic (see Table 1 for a detailed description of each child engagement). Kappas range from .70 to .87.

Global classroom environmental quality was assessed using the Early Childhood Education Rating System–Revised (ECERS–R; Harms, Clifford, & Cryer, 1998) and the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2004). The ECERS–R assesses aspects of the preschool classroom, such as room arrangement and furnishings and relaxation. Scores from the Multi-State study reflect the mean of a fall and a spring observation, and scores from the SWEEP study reflect a single observation day in the fall. Composite ECERS–R scores range from 1 (*inadequate quality*) to 7 (*excellent quality*). Factor analysis of the ECERS–R yielded two factors, one of which was used in the current study: the Teaching and Interactions factor, which consisted of items such as Encouraging Children to Communicate, Discipline, and Interactions among Children. The CLASS rates the emotional climate, classroom climate, and instructional supports for learning in early childhood classrooms. Classrooms were rated from 1 (*low*) to 7 (*high*) on nine dimensions, such as Positive Climate, Teacher Sensitivity, and Concept Development. Data collectors rated the classroom and teacher on the nine dimensions about every 30 min throughout an observation day. The CLASS scores for children in the Multi-State study reflect the mean of a fall and a spring observation, and for children in the SWEEP study a single observation day in the spring. Factor analysis of the CLASS yielded two factors, one of which was used in the current study: Instructional Climate, which is a composite of Concept Development and Quality of Feedback.

Centralized training of all observers was conducted for the three classroom observations. Prior to data collection, observers were certified via reliability tests that compared observers' ratings with experts' ratings. Interrater agreement was evaluated using weighted Kappas, which adjust for agreement due to chance; observers with an overall kappa of at least .60 were certified (see Pianta et al., 2005, for complete details). Kappas ranged from .65 to .81. for the ECERS–R, CLASS, and Snapshot; Kappas of .65 or higher indicate good agreement (Landis & Koch, 1977).

Indicators of School Readiness

Direct child assessments. In the fall and spring, children's language, preliteracy, and mathematics skills were conducted by a different data collector

Table 1
Definitions for the Emerging Academics Snapshot

Code	Description
Activity settings	
Basics	Napping, toileting, standing in line, cleaning up, or waiting between activities
Free choice	Child selects what and where to play or learn, engaging in activities such as individual art projects, blocks, pretend play, and reading
Individual time	Child and the rest of the class each work on a project independently, such as a worksheet or on the computer. The teacher moves around to help
Meals	Eating lunch, breakfast, or snacks
Small group	Small-group activities that are teacher organized or teacher led, such as group art projects, writing stories, or collective building
Whole group	Whole-group activities that are teacher initiated, such as stories, songs, calendar, discussions, book reading, and demonstrations
Preacademic and academic activities	
Esthetics	Art, drama, or music activities
Fine motor	Stringing beads, building with Legos, cutting, or using crayons and markers
Gross motor	Running, skipping, jumping, swinging, riding bikes, or playing games such as basketball, catch, run and chase, dancing, or musical chairs
Letter and sound	With guidance from teacher, child identifies letters, sounds out words, talks about letter-sound relationships, recognizes sounds using rhymes
Mathematics	Rote counting, counting with one-to-one correspondence, skip counting, matching numbers to pictures, making graphs, or playing counting games
Oral language	Child interacts with teacher or peers in talking about stories, or telling development stories of their own, or answering and asking open-ended questions
Prereading	Reading stories, identifying words, recognizing symbols and pictures as having meaning, or practicing a class poem
Read to	Teacher reads books and stories to child, engages in talking about the author, showing the cover, or asking questions about the book
Science	Child explores natural phenomena in their environment, uses science equipment, and reads books or talks about animals, body parts, and life cycles
Social studies	Talking, reading, or engaging in activities about their world (e.g., their neighborhood, their school, the farm, the community workers)
Writing	Writing (or using a computer keyboard to write) numbers or letters
Teacher–child interactions	
Routine	Teacher engages in routine caregiving (e.g., wipe child's nose)
Minimal	Teacher answers child's direct requests for help, or gives simple verbal directives with no reply encouraged, such as "okay," "stop that!"
Simple	Teacher uses some warm or helpful physical contact or verbally answers the child's verbal bids but does not elaborate
Elaborated	Teacher engages in some physical response (e.g., thumbs up, high fives, frown, glare) or acknowledges the child's statements and responds
Scaffold	Teacher (or a more capable peer) does one-on-one work with child and builds on child's initiations, using visuals, concrete objects, and gestures to help child learn. Teacher elicits responses and helps child expand his or her thoughts
Didactic	Teacher lectures, gives instructions, models, asks close-ended questions, or demonstrates, such as counting or saying the days of the week

from the one who conducted the classroom observations.

Children who spoke a home language other than English were given an English language screener, the Pre-LAS2000 (Duncan & De Avila, 1998). The Pre-LAS was first administered in the fall, and children who did not pass in the fall were assessed again in the spring (fall: $\alpha = .89$; spring: $\alpha = .89$). Children who passed the screener (a score of 31 or higher of 40) were administered the English assessment battery (fall: $n = 147$; spring: $n = 117$). Children who did not pass the screener and who spoke Spanish at home were administered the Spanish assessment battery (fall:

$n = 394$; spring: $n = 317$). Children who did not pass the screener and who spoke a language other than Spanish at home were not assessed in the Multi-State but were assessed in English in the SWEEP (fall: $n = 39$; spring: $n = 11$). With the exception of the Peabody Picture Vocabulary Test (PPVT) and the Test de Vocabulario en Imagenes Peabody (TVIP), English and Spanish assessments were collapsed for analysis so that children who were administered the fall assessments in Spanish and the spring assessments in English (i.e., the 117 children who passed the Pre-LAS in the spring) could be retained for analysis.

The PPVT–3rd Edition (PPVT–III; Dunn & Dunn, 1997) is a test of English receptive vocabulary where the child was shown a set of four pictures at a time and asked to select the picture that best represents the word spoken by the examiner (fall: $\alpha = .96$; spring: $\alpha = .96$). A standardized score was computed for this scale. The TVIP (Dunn, Lugo, Padilla, & Dunn, 1986) is the Spanish version of the PPVT (fall: $\alpha = .92$; spring: $\alpha = .93$).

The Oral and Written Language Scale–Oral Expression Scale (OWLS; Carrow-Woolfolk, 1995) assesses children's understanding and use of spoken language (English only; fall: $\alpha = .92$; spring: $\alpha = .91$). After receiving the examiner's verbal stimulus, the child looked at a picture board and responded orally by answering a question, completing a sentence, or generating a sentence. A standardized score was computed for this scale.

Woodcock-Johnson (WJ) III Tests of Achievement (Woodcock, McGrew, & Mather, 2001) are well-established measures of academic achievement, and the Spanish versions are the Batería Woodcock-Muñoz-Revisada: Pruebas de Aprovechamiento (Woodcock & Munoz-Sandoval, 1996). The Applied Problems subtest is a mathematics test that assessed children's ability to comprehend the nature of a problem, identify relevant information, and perform simple calculations (English fall: $\alpha = .84$; English spring: $\alpha = .83$; Spanish fall: $\alpha = .81$; Spanish spring: $\alpha = .79$). The Letter–Word Identification subtest, administered in the SWEEP study only ($n = 1805$), asks children to identify letters and then words (English fall: $\alpha = .89$ English spring: $\alpha = .83$; Spanish fall: $\alpha = .65$; Spanish spring: $\alpha = .89$). Most children moved beyond identifying letters to identifying words—that is, their ceiling set included word identification.

Children's ability to identify letters, numbers, and colors, to count, and to write their names were assessed in both the English and Spanish assessment batteries. For identifying letters, children were shown a set of mixed capital and lowercase letters and asked to identify as many as they could; the maximum score was 26. For identifying numbers, children were shown a sheet of numbers from 1 to 10, presented in random order, and asked to identify as many as they could; the maximum score was 10. For naming colors, children were presented with a page of 10 different colored bears and asked to name the colors; the maximum score was 10. Alphas for identifying letters, numbers and colors for the English and Spanish assessments ranged from .81 to .97. For the counting task, children were asked to count and point with one-to-one correspondence using a picture card with 20 teddy bears. If the child counted to 20 correctly, another sheet of 20 teddy bears was presented to allow continued counting. Finally, children wrote their names and the percent of their name written legibly was calculated.

Teacher report. Teacher report of children's language and literacy skills in the fall and spring was an average of nine items from the teacher questionnaire of the Early Childhood Longitudinal Studies–Kindergarten Cohort (West, Denton, & Germino-Hausken, 2000). The items assessed children's proficiency in speaking (e.g., using complex sentence structure), listening, early reading (e.g., predicting what will happen next), and early writing (e.g., using initial consonants to spell words). The items were rated on a scale from 1 (*not yet*) to 5 (*proficient*; fall: $\alpha = .91$; spring: $\alpha = .92$).

Poverty Status

Information was obtained on whether the children's family lived below the federal poverty threshold ($17,960 for a family of four in 2001; $18,660 for the same family in 2003).

Sociodemographic Covariates

The analysis of group differences across school readiness used the following covariates: child gender, child ethnicity, household size, poverty status, mothers' years of education, and child age at the spring assessment.

Results

Descriptives

The proportion of a day a child spent in a particular activity or setting was computed as the ratio of the number of intervals the activity or setting was observed divided by the total number of intervals the child was observed. Mean proportions of a day children spent in each activity setting, teacher–child interaction, and preacademic activity are summarized in Table 2.

In terms of activity settings, children spent the largest amount of time in free-choice (30%) and whole-group activities (27%), and the least amount of time in individual time (7%) and meals (7%). In terms of preacademic activities, children spent the most time on esthetics (15%), social studies (15%), and science (11%); the least amount of time was spent in prereading (3%), letter–sound (4%), and being read to (5%). In terms of teacher–child interactions, children spent the largest amount of time in didactic interactions (31%) and the second largest amount of time in scaffolding (9%); the least amount of time was spent in routine interactions (1%).

Some of the school readiness indicators were correlated. In language and literacy, the following pairs of outcomes had a correlation coefficient of .45 or higher: naming letters and WJ Letter–Word ($r = .66$; all results reported as significant are $p < .05$ or better), naming letters and percent name written legibly ($r = .45$), naming letters and teacher report of language and literacy ($r = .49$), and OWLS and PPVT ($r = .68$). In mathematics, the following indicator pairs had a correlation coefficient of .45 or higher: WJ applied problems and naming numbers ($r = .55$), WJ applied problems and highest number counted ($r = .49$), and naming numbers and highest number counted ($r = .55$).

Table 2
Descriptives of Snapshot Variables by Snapshot Profiles

	Overall sample		Free play profile (Profile 1)		Individual instruction profile (Profile 2)		Group instruction profile (Profile 3)		Scaffolded learning profile (Profile 4)		Significant differences $(p < .05)$[a]
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Activity settings											
Basics	0.21	0.10	0.20	0.08	0.25	0.09	0.26	0.10	0.16	0.08	SL < FP < II = GI
Free choice	0.30	0.17	0.41	0.12	0.13	0.11	0.15	0.10	0.29	0.13	II = GI < SL < FP
Individual time	0.04	0.07	0.02	0.04	0.21	0.08	0.03	0.04	0.03	0.04	FP = GI = SL < II
Meals	0.12	0.07	0.14	0.07	0.11	0.07	0.10	0.06	0.12	0.07	GI < SL < FP; II < FP
Small group	0.06	0.09	0.04	0.06	0.04	0.07	0.11	0.12	0.06	0.08	FP = II < SL < GI
Whole group	0.27	0.13	0.20	0.09	0.27	0.11	0.36	0.12	0.34	0.12	FP < II < FP < GI
Preacademic and academic activities											
Esthetics	0.15	0.09	0.14	0.09	0.17	0.09	0.15	0.09	0.16	0.10	FP < SL; FP < II; GI < II
Fine motor	0.10	0.08	0.09	0.07	0.17	0.08	0.09	0.07	0.10	0.08	FP = GI = SL < II
Gross motor	0.06	0.06	0.08	0.06	0.03	0.04	0.04	0.05	0.06	0.05	II = GI < SL < FP
Letter and sound	0.04	0.05	0.03	0.03	0.08	0.07	0.05	0.05	0.05	0.05	FP < GI = SL < II
Mathematics	0.08	0.06	0.06	0.05	0.10	0.08	0.10	0.07	0.11	0.07	FP < GI < SL; FP < II
Oral language development	0.06	0.06	0.04	0.04	0.06	0.05	0.05	0.04	0.14	0.08	FP < GI = II < SL
Prereading	0.03	0.04	0.03	0.04	0.04	0.05	0.03	0.04	0.05	0.05	FP = GI < II < SL
Read to	0.05	0.05	0.04	0.04	0.06	0.05	0.06	0.05	0.08	0.06	FP < II = GI < SL
Science	0.11	0.09	0.11	0.09	0.09	0.08	0.09	0.08	0.16	0.12	GI < FP < SL; II < SL
Social studies	0.15	0.11	0.17	0.11	0.08	0.08	0.11	0.08	0.21	0.12	II < GI < FP < SL
Writing	0.01	0.03	0.01	0.02	0.03	0.05	0.01	0.02	0.02	0.03	FP = GI < SL < II
Teacher–child interactions											
Routine	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	None
Minimal	0.03	0.03	0.03	0.03	0.02	0.03	0.02	0.03	0.02	0.03	None
Simple	0.05	0.04	0.06	0.05	0.04	0.04	0.04	0.04	0.05	0.05	II = GI < SL < FP
Elaborated	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.03	0.06	0.05	II = GI < FP < SL
Scaffold	0.09	0.08	0.06	0.06	0.09	0.07	0.08	0.06	0.21	0.08	FP < GI < II < SL
Didactic	0.31	0.16	0.25	0.14	0.39	0.16	0.35	0.14	0.37	0.17	FP < GI < SL < II

Note. FP = free play profile; II = individual instruction profile; GI = group instruction profile; SL = scaffolded learning profile.
[a]Bonferroni post hoc contrasts were used.

Latent Class Analysis

Latent class analysis was conducted in a series of steps, the first of which was fitting a one-profile model to the data to establish a baseline. Models with successively increasing numbers of profiles (up to a model with five profiles, which failed to converge) were then tested. Models were compared using the following model fit indices: the Akaike information criterion (AIC; Akaike, 1987), the Bayesian information criterion (BIC; Schwartz, 1978), and the adjusted BIC (ABIC; Sclove, 1987). Smaller values on these indices indicate better fit, but there is no criterion for "good" fit; therefore, these values are only useful for comparing two or more models. Additionally, an entropy approaching 1.0 indicates a clear distinction between profiles (Celeux & Soromenho, 1996). The four-profile model was the best fitting model with the smallest AIC (−186005.275), BIC (−185277.150), and ABIC (−185667.961), and the largest entropy (0.825). The LCA models were tested using Mplus (Version 4.0; Muthén & Muthén, 1998-2006).

Profiles 1, 2, 3, and 4 included 51%, 9%, 27%, and 13% of the children, respectively. Means and standard deviations of classroom engagements for each profile are summarized in Table 2. Using analysis of variance (ANOVA), we identified significant differences in classroom engagement across profiles. Children in Profile 1 spent by far the most time in free choice (41%), compared to other profiles, and less time than all other groups on a variety of preacademic engagements. Profile 1 is therefore labeled the "free play" profile. Children in Profile 1 also spent a little more time engaged in gross motor activities (8%) than any other profile and more time than Profiles 2 and 3 in social studies (17%). Children in Profile 2 spent much more time than any other profile in individual time (21%), fine motor skills (17%), and letter–sound (8%). Thus, Profile 2 is labeled the "individual instruction" profile. Children in Profile 3 spent more time than any other profile in whole group (36%) and small group (11%), and this profile is therefore labeled the "group instruction" profile. Children in Profile 4 spent much more time engaged in scaffolding interactions with teachers (21%) than any other profile, and also the most time in elaborated teacher–child interactions (6%). In addition, Profile 4 spent more time than any other profile on many preacademic activities. Therefore, Profile 4 is labeled the "scaffolded learning" profile. Profile 4 spent more time than Profiles 2 and 3 on free-choice activities (29%).

Using ANOVA, we found that the snapshot profiles differed significantly on some demographic characteristics. An examination of household size, maternal education, poverty status, and ethnicity showed that children in the free play and scaffolded learning profiles appeared to be somewhat more privileged than children in the individual instruction and group instruction profiles (see Table 3). Specifically, the free play and scaffolded learning profiles have smaller households and more years of maternal education than children in the individual and group instruction profiles. Children in the indi-

Table 3
Demographic Information by Snapshot Profiles

	Free play profile		Individual instruction profile		Group instruction profile		Scaffolded learning profile		Significant differences ($p < .05$)[a]
	M or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	
Household size	4.35	1.43	4.81	1.56	4.57	1.45	4.38	1.34	FP = SL < II = GI
Child is poor	0.57	—	0.69	—	0.57	—	0.50	—	FP = SL = GI < II
Mothers years of education	12.77	2.36	12.06	2.58	12.53	2.37	13.00	2.67	II < GI < SL; II < FP
Child age (spring)	5.04	0.32	5.09	0.30	5.06	0.32	5.06	0.31	None
Latino	22%	—	41%	—	29%	—	28%	—	FP < GI < II; SL < II
African American	16%	—	27%	—	24%	—	14%	—	SL = FP < II = GI
Asian American	15%	—	13%	—	10%	—	14%	—	GI < FP
European American	46%	—	19%	—	37%	—	43%	—	II < GI < FP; II < SL
Gender (male)	50%	—	45%	—	49%	—	50%	—	None

Note. FP = free play profile; II = individual instruction profile; GI = group instruction profile; SL = scaffolded learning profile. Percentages are base rates within each profile.
[a]Bonferroni post hoc contrasts were used.

vidual and group instruction profiles were more likely to be Latino or African American than children in the free play or scaffolded learning profiles. Children in the individual instruction profile were also most likely to be poor compared to all other children. No gender differences emerged between profiles.

Predicting Academic Gains Based on Profile Membership

We assessed the predictive validity of the snapshot profiles by examining fall-to-spring gains in school readiness between the four profiles. We used analysis of covariance (ANCOVA) and covaried for household size, poverty status, maternal education, child race/ethnicity, child gender, and child age. Change was assessed by including fall scores as covariates (i.e., predictors) of the spring scores, such that the residual variance represents change from fall to spring.

Table 4 shows the scores for each profile of children and for each outcome. For example, the score of 12.19 on naming letters in the column "Free play profile" indicates that after adjusting for covariates, children in the free play profile were able to name an average of 12 letters. The overall F tests revealed significant differences across profiles for the following indicators of school readiness: naming letters, WJ letter–word identification, percent of name written legibly, teacher report of language and literacy skills, WJ applied problems, naming numbers, and highest number counted. Bonferroni post hoc contrasts were then used to examine pairwise differences between profiles.

The free play profile made the smallest gains across language and literacy and mathematics.

Table 4
Residual Change Scores From Fall to Spring of Prekindergarten, by Snapshot Profiles

	Free play profile		Individual instruction profile		Group instruction profile		Scaffolded learning profile			
	$n = 1398$		$n = 245$		$n = 755$		$n = 353$			
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	F-test, significant differences ($p < .05$)[a]	Partial eta^2
Language and literacy										
Naming letters	12.19	0.17	14.64	0.41	14.21	0.22	13.73	0.34	$F(3, 2391) = 22.76$*** FP < SL = GI = II	0.03
Naming colors	8.96	0.05	9.00	0.12	9.03	0.06	9.09	0.10	$F(3, 2392) = 0.53$	0.00
PPVT (standardized score)	96.81	0.25	96.77	0.65	97.24	0.34	96.83	0.49	$F(3, 2279) = 0.39$	0.00
TVIP (standardized score)	85.76	1.24	84.41	1.48	84.15	1.18	83.94	2.15	$F(3, 235) = 0.37$	0.01
OWLS (Oral and Written Language)	94.82	0.24	94.29	0.65	94.37	0.33	94.63	0.49	$F(3, 2022) = 0.50$	0.00
WJ letter–word identification	342.92	0.74	351.28	1.35	350.21	0.76	346.68	1.14	$F(3, 1548) = 18.76$*** FP < SL = GI = II	0.04
Percent name legible	83.41	0.63	89.32	1.50	88.85	0.83	85.76	1.24	$F(3, 2359) = 10.82$*** FP < GI = II	0.01
Language and literacy (teacher report)	2.94	0.02	3.15	0.06	3.06	0.03	3.09	0.05	$F(3, 1984) = 6.64$*** FP < SL = GI = II	0.01
Mathematics										
WJ applied problems	412.35	0.36	416.50	0.85	413.14	0.47	412.63	0.70	$F(3, 2372) = 6.73$*** FP = SL = GI < II	0.01
Naming numbers	6.40	0.07	6.85	0.17	6.91	0.10	6.85	0.14	$F(3, 2391) = 7.42$*** FP < SL = GI	0.01
Highest number counted	19.22	0.28	22.58	0.66	21.87	0.37	20.99	0.55	$F(3, 2336) = 14.92$*** FP < SL = GI = II	0.02

Note. FP = free play profile; GI = group instruction profile; II = individual instruction profile; OWLS = Oral and Written Language Scale–Oral Expression Scale; PPVT = Peabody Picture Vocabulary Test; SL = scaffolded learning profile; TVIP = Test de Vocabulario en Imagenes Peabody; WJ = Woodcock-Johnson.

[a] Bonferroni post hoc contrasts were used.

***$p < .001$.

Specifically, the free play profile showed less growth than all other profiles in naming letters, WJ letter–word identification, teacher report of language and literacy skills, and number counting. The free play profile also showed less growth than the individual and group instruction profiles in writing their names, and less growth than the group instruction profile and the scaffolded learning profile on number counting. Finally, the free play profile showed less growth than the individual instruction profile on the WJ applied problems. There was only one significant difference among the individual instruction, group instruction, and scaffolded learning profiles: the individual instruction profile made the greatest gains on the WJ applied problems.

Although effect sizes (partial eta^2) were small, the differences in gains between groups are meaningful. For example, children in the free play profile scored 8 points lower on the WJ letter–word identification than children in the individual and group instruction profiles; this is large for differences that emerged over the course of less than 1 year. Additionally, measurement error for assessments involving younger children tend to be larger than that for older children, and larger measurement errors result in smaller effect sizes (Burchinal, 2008).

In summary, the free play profile showed less growth across indicators of language/literacy and mathematics compared to the other three profiles. The individual instruction profile outperformed all other groups on WJ applied problems.

Snapshot Profiles as Distinct from Classroom Environmental Quality

We wanted to show that the snapshot profiles contribute information about children's classroom experiences that is distinct from and adds to information conveyed by classroom ECERS-R scores. Ideally, we wanted to find no mean differences in ECERS-R scores across the four snapshot profiles. However, there were indeed significant differences: The free play profile ($M = 4.09$, $SD = 0.73$) and the scaffolded learning profile ($M = 4.14$, $SD = 0.75$) each had significantly higher scores than the group instruction profile ($M = 3.42$, $SD = 0.74$) and the individual instruction profile ($M = 3.17$, $SD = 0.63$). In spite of finding significant mean differences, the fairly large standard deviations of ECERS–R scores indicates significant overlap in ECERS–R scores across different snapshot profiles; this in turn suggests that the snapshot profiles provide information not already conveyed by the ECERS–R scores.

To further examine whether the Snapshot profiles contain unique information about children's experiences, we repeated the ANCOVA by profile membership, adding the ECERS–R Teaching and Interactions Scale and the CLASS Instructional Climate Scale as covariates in two separate sets of analyses. All analyses yielded nearly identical results to the original analyses, suggesting again that the snapshot profiles captures information about children's experiences unique from that captured by the ECERS–R and CLASS (results are available upon request).

Poverty Status and Children's Classroom Engagement

To test whether the relationship between gains in school readiness and profile membership varied across poverty status, we conducted a 2 (poverty status) × 4 (snapshot profile) ANCOVA, covarying for gender, ethnicity, household size, poverty status, maternal education, and child age. Table 5 shows adjusted mean scores by profile membership and by poverty status; interaction effects are also shown.

The interaction between poverty status and snapshot profile membership was significant for WJ letter–word identification and highest number counted, and there was a marginal effect ($p = .055$) for WJ applied problems.

For WJ letter–word identification, highest number counted, and WJ applied problems, poor children in the free play, group instruction, and scaffolded learning profiles made smaller gains than their nonpoor peers. In the individual instruction profile, however, poor children actually made greater gains than nonpoor children across these measures of school readiness.

Discussion

It is useful to think about classroom engagement in terms of number of minutes children spend on each activity during a typical program day. For example, the most frequent activity setting was free play (30%), which occupied an average of 45 min of a part-day, 2.5 hr program—quite a longtime. In contrast, < 8 min/day was devoted to each of the following literacy activities: prereading (3%), letter–sound (4%), and being read to (5%).

Moving beyond whole-sample averages, there was substantial variation in children's classroom engagement. Using a person-centered approach, we identified four profiles of children with distinct

Table 5

Residual Change Scores From Fall to Spring of Prekindergarten, by Snapshot Profiles and by Poverty Status

	Free play profile[a]		Individual instruction profile[a]		Group instruction profile[a]		Scaffolded learning profile[a]		Profile Memership × Poverty Status
	Nonpoor	Poor	Nonpoor	Poor	Nonpoor	Poor	Nonpoor	Poor	
	n = 604	*n* = 794	*n* = 76	*n* = 169	*n* = 321	*n* = 434	*n* = 178	*n* = 175	*F*-test
Language and literacy									
Naming letters	12.44	12.00	13.60	15.02	14.72	13.82	13.79	13.72	1.93
Naming colors	9.08	8.87	8.73	9.10	9.05	9.01	9.13	9.05	1.56
PPVT (standardized score)	97.20	96.45	97.52	96.11	98.07	96.44	96.93	96.87	0.71
OWLS	95.21	94.44	95.42	93.32	95.19	93.58	95.85	93.25	1.19
WJ letter–word identification	346.31	340.40	349.05	351.70	350.75	349.90	348.46	345.35	3.19*
Percent name legible	84.45	82.57	87.61	89.98	88.59	89.06	85.57	86.06	0.82
Language and literacy skills	3.00	2.90	3.16	3.13	3.15	2.99	3.12	3.07	0.47
Mathematics									
WJ applied problems	412.72	412.09	415.09	416.81	414.56	412.04	414.44	410.98	2.54†
Naming numbers	6.58	6.26	6.67	6.89	7.13	6.74	6.79	6.94	1.43
Highest number counted	20.53	18.21	21.54	22.55	23.81	20.33	22.09	20.18	2.69*

Note. Comparisons were made between poor and nonpoor children, within classes. Tests of interactions on the TVIP was omitted due to small cell sizes. OWLS = Oral and Written Language Scale–Oral Expression Scale; PPVT = Peabody Picture Vocabulary Test; TVIP = Test de Vocabulario en Imagenes Peabody; WJ = Woodcock-Johnson.

[a] Mean values.

†$p < .06$. *$p < .05$.

patterns of classroom engagement: a free play profile for whom class time was dominated by free-choice activities; two instructional profiles that spent a lot of time receiving teacher instruction, whether in an individual or group format; and one scaffolded learning profile that engaged in large amounts of teacher scaffolding.

We had expected a profile of children that received frequent teacher instruction. We did not expect, however, two instructional profiles that were distinct in terms of the format—individual or group—in which they received teacher instruction. Advantages of individual settings include increased attention for teacher directions and greater need for self-regulation while giving children an opportunity to practice skills they already learned (Stright & Supplee, 2002). An advantage of the group activity setting is that children have the opportunity to share ideas and learn from one another. These results suggest that among the teachers who practice the instructional model of early childhood education, some teachers prefer to deliver instruction via individual seatwork while other teachers prefer to deliver instruction via group work.

Next, we found that the free play profile made the smallest fall-to-spring gains across many measures of language and literacy and mathematics compared to the two instructional profiles and the scaffolded learning profile. Recall that we had included a variety of assessments to explore whether different models of early childhood education were best suited for different domains of learning; our results did not support this hypothesis.

Given that the free play profile was by far the largest profile, making up 51% of the sample children, it was discouraging to find that it was also the profile with the smallest gains in child outcomes. Although these results suggest that the instructional and scaffolding models of early childhood education are more beneficial than the free-choice play model, a caveat exists. Recall that the study sample was at higher demographic risk than the national average: A full 58% of the sample is poor, and the average maternal education is only 12.9 years. Recall also the research showing that instructional support was beneficial for at-risk children, but not necessarily for children not at risk (Hamre & Pianta, 2005). It is possible, therefore, that our finding that classroom instruction was more beneficial than free-choice play is partially the result of having a sample of predominantly at-risk children.

Our study results revealed two further points of interest regarding the scaffolded learning profile. First, the scaffolded learning profile spent a large amount of time in free-choice activities (29%): less than the free play profile (41%) but a lot more than either of the instructional profiles (13% and 15%). These results suggest that free play, when accompanied by high-quality scaffolding interactions with teachers, remains a model of classroom engagement that may be conducive to children's learning. Second, it was disappointing that although the scaffolded learning profile had more than twice as many scaffolding interactions with teachers as any other profile, children in this profile did not experience the greatest growth compared to other profiles. Subsequent analysis revealed a possible explanation: Children in the scaffolded learning profile had the highest fall scores across several child outcomes (see Table 6). Recall also that the scaffolded learning profile was the most sociodemographically advantaged of the four profiles with smaller household sizes, a lower likelihood of being poor, and mothers with higher levels of education. As such, these children may have been better prepared for prekindergarten as a result of educational experiences they received in the home. Also, coming from more advantaged families may be directly related to the frequency of scaffolding interactions: These children may experience more scaffolding interactions with parents and therefore may be

Table 6
Mean Fall and Spring Scores for All Indicators of School Readiness, by Snapshot Profiles

	Free play profile				Individual instruction profile			
	Fall		Spring		Fall		Spring	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Language and literacy								
Naming letters	7.30	8.77	11.66	9.57	7.35	8.98	14.19	9.54
Naming colors	8.23	2.98	8.98	2.29	7.59	3.46	8.67	2.47
PPVT (standardized)	94.36	14.75	96.78	14.10	89.99	14.59	91.92	13.71
TVIP (standardized score)	78.76	11.09	82.67	14.86	82.39	13.00	86.49	14.98
OWLS	92.37	12.77	94.53	12.96	87.42	12.15	89.57	12.32
WJ letter–word identification	330.82	27.10	341.34	27.89	339.58	27.08	355.08	32.12
Percent name legible	59.16	37.81	81.89	28.74	66.34	35.94	90.59	20.52
Language and literacy skills	2.32	0.87	2.97	0.96	2.19	0.86	3.03	1.01
Mathematics								
WJ applied problems	400.78	20.88	411.66	18.94	398.99	21.23	413.10	17.69
Naming numbers	4.14	3.90	6.19	3.71	4.16	3.99	6.74	3.44
Highest number counted	13.68	9.97	18.52	11.46	13.86	10.20	21.60	12.89

	Group instruction profile				Scaffolded learning profile			
	Fall		Spring		Fall		Spring	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Language and literacy								
Naming letters	7.42	8.60	13.86	9.51	9.49	9.28	15.05	9.42
Naming colors	7.97	3.14	8.95	2.26	8.27	3.04	9.19	2.04
PPVT (standardized)	92.59	14.78	95.61	14.13	96.49	16.58	97.89	15.53
TVIP (standardized score)	77.77	11.96	83.88	13.89	77.96	15.66	82.17	15.79
OWLS	89.90	13.10	92.29	12.62	93.96	14.21	95.10	13.79
WJ letter–word identification	330.55	24.23	348.25	24.90	335.32	26.12	348.34	25.96
Percent name legible	60.36	36.69	87.67	23.59	68.82	36.25	88.81	22.40
Language and literacy skills	2.16	0.83	2.96	0.95	2.48	1.00	3.21	0.98
Mathematics								
WJ applied problems	399.27	20.82	411.73	18.08	404.27	20.60	414.77	20.96
Naming numbers	4.21	3.89	6.83	3.60	5.03	4.03	7.36	3.45
Highest number counted	14.09	9.71	21.32	12.39	14.52	10.11	21.15	11.72

Note. OWLS = Oral and Written Language Scale–Oral Expression Scale; PPVT = Peabody Picture Vocabulary Test; TVIP = Test de Vocabulario en Imagenes Peabody; WJ = Woodcock-Johnson.

more likely to elicit and sustain scaffolding interactions with teachers.

Note that for the individual instruction profile, the standard errors of many child outcome measures are larger than for other profiles; this is because the individual instruction profile had fewer children than the other three profiles. Had the standard errors been smaller, more differences between the individual instruction profile and the free play profile may have been statistically significant. Therefore, the small size of the individual instruction profile had the effect of making the results more conservative.

The study examined a variety of child outcomes—from higher order language and problem-solving skills to basic letter and number recognition skills—that are important predictors of later elementary school reading and mathematics. The WJ battery of achievement tests are good predictors of later academic achievement (McGrew, 1986). Less obvious, perhaps, is that basic skills such as naming letters and counting are also important for later academic success. Basic literacy skills developed in preschool such as naming letters and writing one's own name partially determine reading development in early elementary school (Lonigan, Burgess, & Anthony, 2000; Storch & Whitehurst, 2002). Likewise, the lack of basic mathematics skills in kindergarten, such as the ability to name numbers, predicted having a mathematics learning disability in third grade (Mazzocco & Thompson, 2005). Further evidence from an early mathematics intervention program suggests that teaching basic number sense to children at risk for having difficulties in mathematics improves later mathematics performance (Griffin, Case, & Siegler, 1994). Although middle-class children may have the opportunity to develop these basic skills at home, low-income children may not; therefore, it is all the more important that low-income children learn these skills at school. Indeed, Delpit (1996) made the case that low-income African American children need to be taught basic skills so that they can translate their oral language fluency and creativity—which they already possess—into a format that is recognized by mainstream society.

The snapshot profiles captured information about children's classroom experiences not captured by the ECERS–R or the CLASS. ANCOVA of children's gains across profiles remained unchanged after the ECERS–R Teaching and Interactions Scale and the CLASS Instructional Climate Scale were separately added as covariates. In addition, the two snapshot profiles that exhibited greater gains across the prekindergarten year—the individual and group instruction profiles—actually had lower average composite ECERS–R scores compared to the other two profiles. That is, children in *lower* quality classroom environments (measured by ECERS–R) experienced classroom engagement (measured by the Snapshot) that was associated with *greater* gains in school readiness. These results suggest that the Snapshot captured aspects of children's classroom experiences not captured by the ECERS–R and CLASS.

Looking across poverty status, we found that children in the individual instruction profile were more likely to be poor than children in any other profile. This is consistent with prior research suggesting that teachers of low-SES children tend to focus on developing skills in mathematics and literacy (Lee & Ginsburg, 2007), which are often practiced in individual instruction settings. That the individual instruction profile had the highest proportion of poor children is fortuitously joined by the finding that poor children in the individual instruction profile were the only ones to make greater gains than their nonpoor peers on indicators of mathematics and literacy. This finding is consistent with the view some scholars have expressed that minority and low-income children stand to gain more from a curriculum focused on developing academic skills (Delpit, 1996).

Limitations

Free-choice activities teach children complex skills, such as problem-solving, making plans, and comprehension (Johnson & Yawkey, 1988). Unfortunately, these types of higher order skills were not fully assessed by the measures included in the Multi-State and SWEEP studies; thus, gains made by children in the free play profile may not have been captured. Nonetheless, assessments, such as the PPVT, OWLS, and WJ do capture more advanced language and cognitive abilities, and the free play profile still made smaller gains on these assessments than other children. Although this study may not have included the most ideal measures for capturing gains made by the free play profile, the results nonetheless suggest that children in the free play profile made smaller gains in many important domains of development.

In all analyses, we treated children's English and Spanish scores on various assessments as equivalent (with the exception of the PPVT and TVIP). Although most children were assessed at both time points in the same language, some children

($n = 117$) were administered the Spanish assessments in the fall and, after passing the Pre-LAS in the spring, were administered the English assessments in the spring. Those children would have been dropped had we not combined the English and Spanish scores. Although it is not ideal to calculate child gains with Spanish assessments at the first time point and English at the second, it is a conservative measure of gains that preserves the number of cases without sacrificing the validity of the results.

Peer interactions were not captured by the snapshot measure. However, children certainly learn through peer interactions, including through peer scaffolding interactions with same-ability classmates (King, Staffieri, & Adelgais, 1998). Also, given that a large amount of time was spent in free play, it would have been instructive to capture whether, during free play, children were engaged in peer interaction and, if so, what type of peer interaction (e.g., parallel play, pretend play). Coding an activity setting simply as free play is somewhat limited because very different types of peer interactions, with different outcomes for children's learning, could all occur during free play. Capturing children's peer engagement would have added a rich dimension to further understanding the relationship between classroom engagement and child outcomes.

Several measures of language and literacy and mathematics were moderately or highly correlated. But because many assessments highlight the different strengths of different snapshot profiles, it is still important to include each measure. For example, although WJ applied problems and naming numbers were highly correlated ($r = .55$), the individual instruction profile excelled (compared to other groups) on WJ applied problems only. Furthermore, previous studies have often examined these outcome measures together (Howes et al., 2008).

Implications for Policy and Practice

This study described patterns of children's engagement in prekindergarten classrooms and explores whether some patterns of engagement bring about greater gains for children than others. The environmental quality literature thus far has encouraged preschool teachers to think about the arrangement of physical space and the materials available within classrooms. Results from the current study might encourage teachers to think about the allocation of children's time to various classroom engagements in a way that benefits children's learning.

Increasingly, policy makers and advocacy organizations (e.g., Trust for Early Education, National Institute for Early Education Research) are calling for ''high-quality'' prekindergarten for every child. These findings remind us that quality can be assessed in multiple ways that yield different quality ratings for the same classrooms. For example, a class might receive a relatively high ECERS–R score for ample materials but have low levels of teacher–child interactions and may therefore not be optimally preparing children for school.

Teachers, educators, and policy makers also need to consider the factors contributing to the lack of high-quality teacher–child interactions. As pressures from federal and state mandates increase, and school districts increase demands for school-ready children, teachers are asked to do increasingly more. They are often trying to fit into a very short morning (a) circle time that includes calendar, weather, counting, good morning greetings, and a story; (b) small-group time to work on a particular skill; (c) one or two meals; (d) trips to the bathroom; and (e) an inside and outside play period. Although full-day programs do offer more hours in the day, many of the extra hours are spent in napping, toileting, and snacks. To our observers, the classrooms often felt pressured, as if teachers were racing to check off the list of things that need to be done. This kind of atmosphere does not give teachers opportunities to interact with children to talk about their lives, their play, or their ideas, nor does it provide opportunities for teachers to be responsive to children's interests and needs. Results from this study suggest that more quality instructional time spent with teachers, and less free play time spent without teacher guidance or scaffolding, would better prepare children for entering school.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Barnett, W. S., Hustedt, J. T., Friedman, A. H., Boyd, J. S., & Ainsworth, P. (2007). *The state of preschool 2007: State preschool yearbook*. New Brunswick, NJ: Rutgers–The State University of New Jersey, National Institute for Early Education Research.

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2000). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Burchinal, M. R. (2008). How measurement error affects the interpretation and understanding of effect sizes. *Child Development Perspectives*, *2*, 178–180.

Burchinal, M. R., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2008). *Child care quality and child outcomes: Multiple-studies analyses*. Paper presented at the University of California, Los Angeles.

Burchinal, M. R., Roberts, J. E., Riggins, R. Jr, Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, *71*, 338–357.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales (OWLS)*. Circle Pines, MN: American Guidance Service.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters. *Journal of Classification*, *13*, 195–212.

Delpit, L. (1996). *Other people's children: Cultural conflict in the classroom*. New York: New Press.

Duncan, S. E., & De Avila, E. (1998). *PreLAS 2000*. Monterey, CA: CTB/McGraw-Hill.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: Pearson's Assessment Group.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de vocabulario en imagenes peabody*. Circle Pines, MN: American Guidance Service.

Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, *76*, 949–967.

Harms, T., Clifford, R., & Cryer, D. (1998). *The Early Childhood Environment Rating Scale* (Rev. ed.). New York: Teachers College Press.

Henderson, S. D., Many, J. E., Wellborn, H. P., & Ward, J. (2002). How scaffolding nurtures the development of young children's literacy repertoire: Insiders' and outsiders' collaborative understandings. *Reading Research and Instruction*, *41*, 309–330.

Howes, C. (2000). Social-emotional classroom climate in child care, child-teacher relationships and children's second grade peer relations. *Social Development*, *9*, 191–204.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs *Early Childhood Research Quarterly*, *23*, 27–50.

Howes, C., & Smith, E. W. (1995). Relations among child care quality, teacher behavior, children's play activities, emotional security, and cognitive activity in child care. *Early Childhood Research Quarterly*, *10*, 381–404.

Johnson, J. E., Christie, J. F., Yawkey, T. D., & Wardle, F. (1987). *Play and early childhood development* (Vol. 269). Glenview, IL: Scott, Foresman.

Johnson, J. E., & Yawkey, T. D. (1988). Play and integration. In T. D. Yawkey (Ed.), *Integrative processes and socialization: Early to middle childhood* (pp. 97–117). Hillsdale, NJ: Erlbaum.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*, 134–152.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lee, J. S., & Ginsburg, H. P. (2007). Preschool teachers' beliefs about appropriate early literacy and mathematics education for low- and middle-socioeconomic status children. *Early Education and Development*, *18*, 111–143.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*, 596–613.

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*, 115–157.

Mazzocco, M. l. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*, *20*, 142–155.

McGrew, K. S. (1986). A review of the differential predictive validity of the Woodcock-Johnson Scholastic Aptitude clusters. *Journal of Psychoeducational Assessment*, *4*, 307–317.

Meyer, L. A., Wardrop, J. L., Hastings, C. N., & Linn, R. L. (1993). Effects of ability and settings on kindergartners' reading performance. *Journal of Educational Research*, *86*, 142–160.

Muthén, L. K., & Muthén, B. O. (1998-2006). *Mplus user's guide* (4th ed.). Los Angeles, CA: Author.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*, 144–159.

Pianta, R., La Paro, K. M., & Hamre, B. K. (2004). *Classroom Assessment Scoring System (CLASS)*. Charlottesville: University of Virginia.

Pianta, R., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *Elementary School Journal*, *102*, 225–238.

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weister, B. (2001). *Emerging academics snapshot*. Los Angeles: University of California at Los Angeles.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343.

Smith, A. B. (1996). The early childhood curriculum from a sociocultural perspective. *Early Child Development and Care*, *115*, 51–64.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Stright, A. D., & Supplee, L. H. (2002). Children's self-regulatory behaviors during teacher-directed seat-work, and small-group instructional contexts. *Journal of Educational Research*, *95*, 235–244.

Tonyan, H. A., & Howes, C. (2003). Exploring patterns in time children spend in a variety of child care activities: Associations with environmental quality, ethnicity, and gender. *Early Childhood Research Quarterly*, *18*, 121–142.

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Trans.). New York: MIT Press and Wiley.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergartners: NCES 2000-070*. Washington, DC: National Center for Educational Statistics.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III: Tests of Achievement*. Itasca, IL: Riverside.

Woodcock, R. W., & Munoz-Sandoval, A. F. (1996). *Batería Woodcock Muñoz Revisada: Pruebas de Aprovechamiento*. Itasca, IL: Riverside.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1550–1564

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement

Katherine C. Pears
Oregon Social Learning Center and Center for Research to Practice

Philip A. Fisher
Oregon Social Learning Center, Center for Research to Practice, and University of Oregon

Jacqueline Bruce
Oregon Social Learning Center and Center for Research to Practice

Hyoun K. Kim and Karen Yoerger
Oregon Social Learning Center

In this study, 85 maltreated foster children and 56 nonmaltreated community children (*M* age = 3–6 years) were assessed across kindergarten and first grade to examine the hypothesis that inhibitory control and caregiver involvement mediate associations between a history of maltreatment and foster placement and early school adjustment. Specifically, academic and social-emotional competence were evaluated. The maltreated foster children performed more poorly in academic and social-emotional competence. Inhibitory control fully mediated the association of maltreatment and foster placement with academic competence, whereas inhibitory control and caregiver involvement mediated their association with social-emotional competence. The results suggest that inhibitory control and caregiver involvement might be promising targets for school readiness interventions for foster preschoolers.

Maltreated foster children often display difficulties in a number of cognitive and psychosocial domains (Leslie et al., 2000; Pears & Fisher, 2005), which might impinge upon success in school. Thus, it is not surprising that many of these children have scholastic difficulties, including high rates of special education placement, academic failure, and school dropout (Blome, 1997; Courtney & Dworksy, 2006; Geenen & Powers, 2006; Goerge, Van Voorhis, Grant, Casey, & Robinson, 1992; Smithgall, Gladden, Howard, Goerge, & Courtney, 2004). However, few researchers have explored how foster children fare at elementary school entry, a critical developmental milestone. In the general population, kindergarten and first-grade academic and social-emotional performance predict trajectories of school and psychosocial adjustment across later elementary and middle school grades (Duncan et al., 2007; Lesaux, Rupp, & Siegel, 2007; McClelland, Acock, & Morrison, 2006; Nagin & Tremblay, 2001). Additionally, successful adjustment in the early school years is linked to educational attainment and avoidance of psychosocial problems in adulthood (Fothergill et al., 2008). Understanding foster children's early school adjustment might be pivotal in preventing negative outcomes.

In the current study, we had two primary goals. First, we sought to explore whether disparities in school adjustment can be observed in maltreated foster children as early as kindergarten and first grade. Second, we sought to identify factors that mediate the association between a history of maltreatment and foster placement and early school adjustment.

Components of Early School Adjustment

Converging evidence suggests that successful school achievement in the early grades requires

---

This research was supported by the following grants: MH059780, NIMH, U.S. PHS; and DA021424, NIDA, U.S. PHS. The authors would like to thank Kristen Greenley for project management, Katie Lewis and Matthew Rabel for editorial assistance, and the children and families who participated in the project.

Correspondence concerning this article should be addressed to Katherine C. Pears, Oregon Social Learning Center, 10 Shelton McMurphey Boulevard, Eugene, OR 97401. Electronic mail may be sent to katherinep@oslc.org.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0017

both academic and social-emotional competence (Blair, 2002; Peth-Pierce, 2000; Raver, 2002). Early academic competence includes literacy skills such as language and reading abilities, which are predictive of later academic achievement (Duncan et al., 2007; NICHD Early Child Care Research Network, 2005). For example, Duncan and colleagues showed that early language and reading abilities positively predict achievement in reading and math up to 9 years later. Additionally, better language skills appear to serve a protective function for children exposed to risk factors that might otherwise lead to poor school adjustment (Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006). Conversely, early academic difficulties, particularly in language and reading, appear to predict trajectories of increasing academic and social-emotional problems across the school years (Chatterji, 2006; Hooper, Roberts, Zeisel, & Poe, 2003).

Social-emotional competence comprises three interrelated sets of skills: prosocial behaviors, emotion regulation, and behavior regulation. Key prosocial behaviors include the abilities to enter peer groups, to share materials with others, to cooperate, and to maintain social interactions. These skills predict positive peer and teacher relationships, which are linked to higher classroom participation and achievement (Ladd, Birch, & Buhs, 1999; Pianta & Stuhlman, 2004). Emotion regulation involves a child's abilities to modulate negative emotions (e.g., frustration) and to control positive emotions (e.g., exuberance) so that they do not disrupt academic performance and social relationships (Blair, 2002; Graziano, Reavis, Keane, & Calkins, 2007). Better emotion regulation is linked to higher academic achievement and better social skills in school within grades (Graziano et al., 2007; Howse, Calkins, Anastopoulos, Keane, & Shelton, 2003) and across grades (Newman, Noel, Chen, & Mastopoulos, 1998; Trentacosta & Izard, 2007). Behavior regulation involves a child's abilities to focus on schoolwork, to avoid disrupting classroom activities, and to make successful transitions between activities. Recent studies have linked greater behavior regulation in kindergarten and first grade to higher academic achievement concurrently (Howse, Lange, Farran, & Boyles, 2003) and longitudinally through middle school (Duncan et al., 2007; McClelland et al., 2006). Although these three components of social-emotional competence are conceptually distinct, they have been shown to be strongly intercorrelated (Graziano et al., 2007; Trentacosta & Izard, 2007).

Maltreatment and Foster Placement and Early School Adjustment

School achievement in foster children has been investigated previously (Blome, 1997; Courtney & Dworksy, 2006; Geenen & Powers, 2006; Goerge et al., 1992; Smithgall et al., 2004). However, these studies were generally concentrated on children in the middle elementary grades and higher. Few studies have been focused on foster children in early elementary school. Additionally, because these studies were generally based on archival record data, most only involved the assessment of academic functioning. A number of researchers have examined academic and social-emotional competence in maltreated children in general (e.g., Anthonysamy & Zimmer-Gembeck, 2007; Kurtz, Gaudin, Wodarski, & Howing, 1993; Veltman & Browne, 2001). This line of research is relevant to foster children because the majority of young foster children come into care due to maltreatment. However, foster children have also experienced multiple caregiver disruptions. Consequently, maltreated foster children might exhibit differences in school adjustment compared to maltreated children in general. Given the importance of adjustment at school entry to later school functioning and psychosocial outcomes (Duncan et al., 2007; Fothergill et al., 2008), understanding whether maltreated foster children show early deficits in academic and social-emotional competence is critical to the development of targeted interventions to improve school outcomes for these children.

Potential Mediational Effects on Early School Adjustment in Maltreated Foster Children

In addition to identifying maltreated foster children's deficits in early school adjustment, understanding the factors that mediate the association between a history of maltreatment and foster placement and school adjustment is essential to preventive intervention efforts. Prior research results have suggested at least two mechanisms that leave maltreated foster children vulnerable to difficulties in school: inhibitory control and caregiver involvement. Inhibitory control, a cognitive process that falls under the rubric of executive functioning, is the capacity to voluntarily inhibit or regulate prepotent attentional or behavioral responses (Casey, Tottenham, & Fossella, 2002). It has an extended developmental trajectory and changes over time from being an externally driven process to an internalized ability (Kochanska, Coy,

& Murray, 2001). Because young children are reliant on their caregivers for self-control, disruptions and inconsistencies in the early caregiving environment, such as maltreatment and foster placement, are likely to have deleterious effects on later inhibitory control. Past research findings have shown that caregiver disruptions have negative effects on inhibitory control in general and foster care populations. In a recent study of a community sample of preschool-aged children, Lengua, Honorado, and Bush (2007) found that the number of household moves predicted inhibitory control abilities. Lewis, Dozier, Ackerman, and Sepulveda (2007) showed that adopted children who had experienced multiple prior foster placements performed significantly worse on an inhibitory control task than nonadopted community children and adopted children who had experienced only one prior foster placement. Additionally, executive functions, particularly inhibitory control, have been linked to harsh parenting and maltreatment (Beers & DeBellis, 2002; Valiente, Lemery-Chalfant, & Reiser, 2007), suggesting that foster children's histories of maltreatment might also impinge on their inhibitory control.

Maltreated foster children's potential deficits in inhibitory control suggest a pathway through which maltreatment and foster placement might affect school outcomes. Prior research findings have shown that early inhibitory control is central to school success. Specifically, better inhibitory control in the preschool period has been linked to better academic and social functioning in first grade (Blair & Razza, 2007; NICHD Early Child Care Research Network, 2003). Additionally, inhibitory control appears to be positively associated with concurrent feelings of school liking in later elementary grades (Valiente, Lemery-Chalfant, & Castro, 2007). Finally, children's inhibitory control appears to partially mediate the association between family environment and children's readiness for kindergarten (NICHD Early Child Care Research Network, 2003), suggesting that inhibitory control might provide a link between early experiences and school outcomes.

Maltreated foster children are likely to be at a disadvantage in terms of another important precursor of success in school: caregiver involvement. Caregiver involvement in schooling (e.g., promotion of literacy and homework activities at home, assistance in organizing school materials, direct involvement in classroom activities, and contact with teachers and school personnel) has been shown to be a significant predictor of academic achievement across studies (Jeynes, 2005). It has been associated with higher grades, better attitudes toward schoolwork, greater participation in classroom learning activities, higher attendance rates, fewer placements into special education, fewer suspensions, and greater educational attainment over time (Barnard, 2004; Christenson, 1999).

The majority of young foster children have experienced abusive and neglectful environments in their families of origin. Although there are no prior studies of the school involvement of abusive parents, the results from one study suggest that parents might be less involved in schooling when the family is experiencing difficult life circumstances (Lamb-Parker et al., 2001). Additionally, once placed into foster care, the children often experience multiple caregiver changes. Pears and Fisher (2005) showed that preschool-aged maltreated foster children had experienced an average of 3.47 transitions. The caregiver changes associated with foster placements are likely to interfere with caregiver involvement in schooling for several reasons. First, the caregiver might not have adequate information about the child's academic history. For example, in one study, information on past academic performance was missing from 53% of foster children's school records (Parrish et al., 2001). Second, the caregiver might not have an opportunity to become involved in schooling if the child is in the home for only a limited period of time. In 2005, half of all children exiting foster care had been in their placements between 1 and 11 months (Child Welfare Information Gateway, 2007). Given the potential barriers to foster caregivers' involvement in schooling, it is not surprising that 65% of foster children in high school reported that their caregivers had never attended a teacher conference, 73% reported that their caregivers had never visited their class, and 70% reported that their caregivers had never volunteered at school (Blome, 1997). Although a limited number of studies have explored factors related to caregiver involvement in this population, caregiver involvement at school entry and its role as a mediator in the association between maltreatment and foster placement and early school adjustment have not yet been examined.

Goals of the Study

As noted earlier, the purpose of this study was two-fold. Our first goal was to explore whether deficits in school adjustment can be observed as early as kindergarten and first grade in maltreated foster children. To accomplish this, maltreated foster children were compared to nonmaltreated children living with their biological parents. We hypothesized that the maltreated foster children would exhibit lower scores on academic and social-emotional competence than their nonmaltreated peers. The second goal was to use longitudinal data to test the hypotheses that the pathways between a history of maltreatment and foster placement and poor early school adjustment (i.e., low academic and social-emotional competence) would be mediated by the following: child inhibitory control during preschool and caregiver involvement at kindergarten entry. Potential confounding variables, including the children's age and general cognitive ability, were controlled in the path model. School characteristics (e.g., socioeconomic status [SES] composition and class size) can affect school adjustment (Konstantopoulos, 2006) and were thus included in the model as control variables. Although little is known about the school contexts of foster children, one group of researchers found that foster children were less likely to attend magnet schools and were disproportionately concentrated in the lowest achieving schools (Smithgall et al., 2004).

Method

Participants

The participants in this study were a subsample of children from an efficacy trial to evaluate a treatment foster care program for preschool-aged children. The entire sample included a foster care (FC) group of 117 maltreated foster children and a community comparison (CC) group of 60 age- and SES-matched, nonmaltreated children living with their biological families. For the FC group, children between the ages of 3 and 6 entering new foster placements were referred to the study through the local child welfare system. The inclusion criteria for the FC group specified placement in a nonrelative foster home, an expectation of remaining in foster care for at least 3 months, and English as the primary language of the child and foster caregiver. A staff member contacted each child's caseworker (the legal guardian while the child is in care) and requested consent for the child to participate in the study. Upon receiving caseworker consent, a staff member then contacted the foster caregiver to invite him or her to participate. For a child to be successfully recruited, the caseworker and foster caregiver had to consent to participate.

The CC group was recruited via flyers posted at local supermarkets, day-care centers, and Head Start classrooms and through advertisements in local newspapers and newsletters. The CC families were included in the study if the child had lived consistently with at least one biological parent, household income was no more than $30,000, parental education level was less than a 4-year college degree, and the family had no prior involvement with the child welfare system.

The FC and CC children were 3–6 years old when they were recruited to participate in the study for a period of 24 months. Because the mean age of the children at the baseline assessment was 4.46 years ($SD = 0.81$), most children were not in school at study entry. Data on performance in kindergarten and first grade were collected for 85 FC and 56 CC children who were still participating in the study when they entered those grades. Of the 36 children (32 FC and 4 CC) for whom school data were not available, 28 (26 FC and 2 CC) discontinued their participation in the study before they entered school, 5 were too young to have entered school by the end of the 24-month assessment period (4 FC and 1 CC), 2 were homeschooled (1 FC and 1 CC), and 1 FC child was in second grade by the time that school data were collected. Children for whom school data were not available did not differ significantly on age, sex, ethnicity, or general cognitive functioning from those for whom the data were available.

Although all 85 FC children were in foster care at study entry, 12 children had been permanently placed with biological relatives, and 8 children had been adopted by nonrelatives when the school assessments were conducted. There were no differences between these children and the other FC children on any of the study variables; thus, all children were retained in the FC group. As noted earlier, all FC children were part of a larger efficacy trial to evaluate a treatment foster care program (Multidimensional Treatment Foster Care for Preschoolers; Fisher, Gunnar, Chamberlain, & Reid, 2000) and were assigned to intervention or comparison groups at study entry. (The CC children were not randomized to different conditions.) Maltreated foster children from the intervention ($n = 45$) and control ($n = 40$) groups were combined in the present sample because there were no mean differences between the two groups on any of the variables used in the present study (either the individual scores or the composites) and because correlational analyses with treatment status as a covariate did not reveal any experimental effects.

The FC and CC groups did not differ on mean child age, gender, or ethnicity. Mean age at the first school assessment was 5.89 years ($SD = 0.47$) in the

FC group and 5.81 years (*SD* = 0.49) in the CC group. Boys comprised 53% of the FC group and 54% of the CC group. The ethnicity breakdown of the FC group was: 85% European American, 1% African American, 8% Latino, and 6% Native American. The ethnicity breakdown of the CC group was: 79% European American, 5% African American, 7% Latino, 7% Native American, and 2% Pacific Islander. The study included 22 sibling dyads (12 FC and 10 CC) and 3 sibling tetrads (3 FC).

The FC children had experienced an average of 1.84 placements, 3.73 transitions between placements, and 164 days in care prior to study entry. All FC children had histories of maltreatment. Child welfare case records coded using the Maltreatment Classification System (Barnett, Manly, & Cicchetti, 1993) revealed that the children had experienced physical abuse (36%), sexual abuse (27%), physical neglect (81%), supervisory neglect (87%), and emotional maltreatment (89%). No CC children had been reported for child maltreatment by study entry (as verified by a search of child protective service records).

Procedure

As noted earlier, the children entered the study between 3 and 6 years of age and participated in the study for 24 months. If the children entered school during the time that they participated in the study, school data were collected in the fall and spring of their kindergarten year and in the spring of their first-grade year. Some of the children had attended kindergarten and first grade by the end of the 24 months and participated in all school assessments. Some children, however, were young enough at study entry that they participated only in the kindergarten assessments, whereas some children were old enough at study entry that they participated only in the kindergarten spring assessment, the first-grade assessment, or both. Thus, not every child had data for all school assessment points. One hundred twenty-five children (89%) had data for the kindergarten fall assessment, 107 (76%) had data for the kindergarten spring assessment, and 96 (68%) had data for the first-grade spring assessment. Sixty-nine children (49%) participated at all three assessment points, 49 children (35%) participated at two assessment points, and 23 children (16%) participated at one assessment point. To maximize the number of children included in the analyses and the data available for each child, measures were combined across the kindergarten and first-grade assessments.

In addition to the school assessments, all children and their families completed laboratory-based measures and questionnaires at 3-month intervals over the 24 months of the study. At the baseline assessment (study entry), the children completed laboratory-based measures to assess inhibitory control and general cognitive functioning. At the 3-month assessment, the caregivers completed a questionnaire measure of inhibitory control. Because the children entered the study at different ages, the time between the baseline assessment and the first school assessment varied (*M* = 17 months, *SD* = 9 months).

For the current study, individual scales and multimethod, multiagent composites were created according to Patterson and Bank's (1986) method and were examined in analyses. All scales and composites had to show an internal reliability of .60 or higher (standardized alpha), and all items in a scale or composite had to show an item–total correlation of .20 or higher.

Group Status

In the current study, a history of maltreatment and foster placement was represented by a dichotomous variable. The FC children were coded as 1 (*history of maltreatment and foster placement*), whereas the CC children were coded as 0 (*no history of maltreatment or foster placement*).

School Adjustment

Academic Competence

This aspect of school adjustment was assessed via a composite measure of teacher report and school records data. Teachers rated the children's performance in language, language comprehension, articulation, and self-help abilities on a scale of 1 (*superior*) to 5 (*delayed*) as a measure of the children's language and learning skills. The scores were reverse coded: Higher scores indicated higher skill. The language and learning skills scale showed good internal consistency at all school assessments. Reliability information for all school adjustment measures across time is shown in Table 1. Additionally, we collected information from school records on the number of special academic services received during kindergarten and first grade. These special services were classified according to the following categories: special education in the general classroom, special education in a self-contained classroom, and Chapter I/Title I services. The number

Table 1
Reliability Statistics for Subscales of the School Adjustment Measures

	Kindergarten fall	Kindergarten spring	First-grade spring
Language and learning skills	.87	.90	.84
Prosocial behavior	.90	.89	.91
Emotion regulation (r values)	.41	.48	.60
Behavior regulation	.82	.78	.79

Note. All values represent standardized alpha coefficients unless otherwise noted.

of special academic services was reverse scored. The language and learning skills scales were combined across time points ($\alpha = .77$), as were the numbers of special academic services received ($r = .35$, $p < .05$). The resulting language and learning skills and number of academic services scores were significantly correlated ($r = .51$, $p < .001$), and these scores were standardized and combined to produce the academic competence composite score. Higher values represent better academic competence.

Social-Emotional Competence

This composite score was composed of three domain scores—prosocial behaviors, emotion regulation, and behavior regulation—and was assessed using four teacher-report measures and one child-report measure. The Teacher Social Competence Scale (Conduct Problems Prevention Research Group, 1995) is a 12-item measure that assesses children's prosocial behavior and communication skills and emotion regulation. The Loneliness and Social Dissatisfaction Questionnaire for Young Children–Teacher Version (Cassidy & Asher, 1992) is a 12-item instrument designed to assess child behavior on four dimensions: aggression, disruptiveness, prosocial behavior, and withdrawal. The Walker-McConnell Scales of Social Competence and School Adjustment (Walker & McConnell, 1988) is a 45-item teacher-report instrument measuring teacher-preferred behaviors, peer-preferred behaviors, and school adjustment. The teachers also completed the Emotion Regulation Checklist (Shields & Cicchetti, 1997), a 27-item questionnaire that measures emotion regulation and lability and negativity. The children completed the Seattle Personality Questionnaire (Greenberg & Lengua, 1995), a 44-item measure that forms eight subscales, including the Conduct Problems subscale.

Prosocial behavior. This score comprised three scales: the six-item Prosocial Behaviors/Communication Skills subscale of the Teacher Social Competence Scale, the three-item Prosocial Behaviors subscale of the Loneliness and Social Dissatisfaction Questionnaire for Young Children–Teacher Version, and the 17-item Peer-Preferred Behaviors subscale of the Walker-McConnell Scales of Social Competence and School Adjustment. At each time point, the individual scales were standardized and averaged to form the prosocial behavior domain score (see Table 1). These scores were then averaged across time points ($\alpha = .78$).

Emotion regulation. This score comprised two scales: the six-item Emotion Regulation subscale of the Teacher Social Competence Scale and the eight-item Emotion Regulation subscale of the Emotion Regulation Checklist. The scales were significantly positively correlated at each time point (see Table 1). Thus, the scales were standardized and averaged at each time point and then averaged across time points ($\alpha = .80$).

Behavior regulation. This score comprised four subscales: the three-item Aggression and three-item Disruptiveness subscales of the Loneliness and Social Dissatisfaction Questionnaire for Young Children–Teacher Version, the 15-item Lability/Negativity subscale of the Emotion Regulation Checklist, and the eight-item Conduct Problems subscale of the Seattle Personality Questionnaire. Within each time point, the scales were standardized and averaged, producing internally consistent domain scores (see Table 1). These scores, which were averaged across time points ($\alpha = .83$), were reverse scored: Higher scores reflected better behavior regulation.

The goal in producing separate social-emotional domain scores was to use the individual scores in the analyses. However, in preliminary analyses, the domain scores showed strong associations with one another ($rs = .75$–$.92$). To avoid problems of multicollinearity in the analyses, the prosocial behaviors, emotion regulation, and behavior regulation domain scores were averaged to produce an overall social-emotional competence composite score. The composite showed high internal consistency ($\alpha = .96$).

Mediators of the Effects of Maltreatment and Foster Placement on School Adjustment

Inhibitory Control

Inhibitory control was assessed via laboratory child assessments and caregiver reports. The laboratory assessments included the modified Stroop task (Gerstadt, Hong, & Diamond, 1994), which is associated with other measures of inhibitory control (Carlson & Moses, 2001). The children were instructed to say "night" whenever they saw a card with a yellow sun on a white background and to say "day" whenever they saw a card with a white moon on a black background. Sixteen cards were presented in a pseudorandom order. The mean correct score was used in analyses. Children also completed the Dimensional Change Card Sort task (adapted from Frye, Zelazo, & Palfai, 1995), which correlates well with other inhibitory control measures (Carlson & Moses, 2001). In this task, the children were asked to sort a series of cards first according to one criterion and then according to another criterion, which required inhibition of the tendency to respond according to the first criterion. For three of the postswitch trials (after the sorting criterion changed), the cards were incompatible with the preswitch rules. The mean correct score for the incompatible trials was used in analyses. The attention/executive function domain of the NEPSY (Korkman, Kirk, & Kemp, 1998) was also used. For 3- and 4-year-olds, this domain includes two subscales: Visual Attention and Statue. (The NEPSY was not administered to children who were over the age of 5 at the baseline assessment [$n = 40$]; thus, this score was not included in the composite for these children.) The caregiver reports were assessed using the Inhibitory Control and Attention Focusing scales of the Children's Behavior Questionnaire (Rothbart, Ahadi, Hershey, & Fisher, 2001), which assess behavior and attention regulation, respectively. The laboratory measures and caregiver ratings were standardized and averaged to form the inhibitory control composite ($\alpha = .68$).

Caregiver Involvement in Schooling

Caregiver involvement in schooling was assessed in the fall of the kindergarten year via a nine-item, teacher-report scale measuring caregiver involvement in preparing children for school, the frequency of caregiver communication with the teacher, and caregiver interest in solving problems at school. The scale showed good internal consistency ($\alpha = .89$).

Covariates

General Cognitive Functioning

At the baseline assessment, the children completed the Vocabulary and Block Design subtests of the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI–R; Wechsler, 1989). These subtests are considered to be the strongest measures of verbal and performance intelligence, respectively, and are highly correlated with other general intelligence measures (Wechsler, 1989). The scaled scores for these subtests were averaged. In the general population, the mean of WPPSI–R scaled scores is 10 ($SD = 3$).

Age at the Baseline Assessment

Because inhibitory control is associated with age (Kochanska, Murray, & Harlan, 2000), we controlled for age at the baseline assessment in all analyses. Additionally, as noted earlier, the time between the baseline assessment and the first school assessment varied. However, child age at the baseline assessment was strongly correlated with time between the baseline and school assessments ($r = -.81$, $p < .001$). Rather than adding another control variable, we used child age at the baseline assessment to control for time between the baseline and school assessments.

School Characteristics

Indices of school characteristics were retrieved from the Oregon State Department of Education's Web site (e.g., number of children receiving free or reduced lunches, average class size, and percentage of teachers with a master's degrees or higher; Oregon Department of Education, 2007). These indices were averaged across kindergarten and first grade. These data were not available at either kindergarten or first grade for 7 children: 3 children in private schools (1 FC and 2 CC), 3 FC children in specialized schools, and 1 CC child living out of state. Additionally, each child's school was classified as a regular public school, a private or magnet school, or a specialized treatment school. In a few cases, a child attended a regular school for 1 year and a private or magnet school the other year; in such instances, the child was coded as attending a private or magnet school. No children had been in a private or magnet or regular school and a specialized treatment school.

Data Analysis Plan

Preliminary analyses were conducted to examine patterns of normality and missing data. Next, differences between the FC and CC groups on the hypothesized mediators, covariates, and outcomes were examined using Hotelling's T^2 tests to control for inflated error rates. This was followed by tests of the associations between the hypothesized mediators, covariates, and outcomes using bivariate correlations. Finally, the full path model with the hypothesized mediators, covariates, and outcomes was tested using Mplus (Version 5.1; Muthén & Muthén, 2007).

Results

Preliminary Analyses

Prior to the analyses, all variables were examined for significant deviations from normality and outliers. The caregiver involvement variable was significantly skewed and was log-transformed. Additionally, for the academic and social competence variables, there was one extreme outlier for each (±3 *SD* of the mean). These scores were recoded to the next value that was within 3 *SD* of the sample mean. In general, for the composite scores and other variables, there were few missing data (< 6%), with the exception of caregiver involvement. Because caregiver involvement in the fall of the kindergarten year was used, missed assessments at this time point resulted in data missing for 17% of the sample on this variable. In the path analyses using Mplus, missing data were estimated using full information maximum likelihood techniques. To account for the nonindependence of the sibling groups, the Hotelling's T^2 tests and bivariate associations were repeated with randomly selected siblings omitted. There were no significant changes in the patterns of results. Additionally, the cluster design approach in Mplus, which adjusts standard errors for the nonindependence of data, was used for the path analyses. Thus, to maximize power, all siblings were retained in the analyses. Finally, to determine if there were gender differences on the hypothesized mediators, covariates, and outcomes, gender was entered into the Hotelling's T^2 tests. The overall T^2 values were not significant. Thus, gender was not included as a covariate.

Group Differences

To control for the possibility of inflated errors with multiple single *t* tests, two Hotelling's T^2 tests were used to examine group differences. The hypothesized mediators, covariates, and outcomes were tested in two groups: one included the child and family characteristics (mediators, general cognitive ability, and school adjustment) and the other included the school characteristics variables. The overall T^2 statistic (adjusted for unequal covariances) for the child and family characteristics indicated that there were significant group differences, $T^2(5, 103) = 33.51$, $p < .01$. As shown in Table 2, follow-up individual *t* tests indicated that the groups differed on all variables. (Negative values on composite scores indicate that the scores were below the sample mean.)

Table 2

Means and Standard Deviations by Groups for the Hypothesized Mediators, Covariates, and Outcomes

	Foster care group		Community comparison group		
	M	*SD*	*M*	*SD*	*t*
School outcome measures					
Academic competence	−0.16	0.92	0.30	0.67	2.93**
Social-emotional competence	−0.21	0.70	0.25	0.73	3.42**
Hypothesized mediators					
Inhibitory control	−0.11	0.69	0.28	0.53	3.28**
Caregiver involvement	−0.20	0.15	−0.14	0.12	2.50**
Covariates					
Percentage of students receiving free or reduced lunch	52.51	16.89	46.60	17.05	1.96†
Average class size	23.31	2.28	23.33	1.53	0.07
Percentage of teachers with a master's degree or higher	45.87	13.10	52.94	13.99	2.96**
General cognitive ability	8.02	2.14	9.80	2.06	4.42**
Age at the baseline assessment	4.54	0.82	4.33	0.79	1.51

Note. Negative values on composite scores indicate that the scores were below the sample mean.
†$p < .10$. **$p < .01$.

For the school characteristics, there was a significant overall T^2 statistic (adjusted for unequal covariances), $T^2(3, 106) = 12.36$, $p < .01$. Examination of the individual t tests revealed that there was a significant group difference in the percentage of teaching staff with a master's level degree or higher (Table 2).

Finally, a chi-square test was used to determine if there were group differences in the likelihood of attending private or magnet schools. In the current study, 25% of the CC children attended a private or magnet school compared to 6% of the FC children, $\chi^2(1) = 10.58$, $p < .01$. (The group difference remained even when the 5 FC children in treatment schools were excluded from analyses.)

Correlational Analyses

The bivariate correlations among the variables are displayed in Table 3. Consistent with the group differences, a history of maltreatment and foster placement was associated with lower academic and social-emotional competence in kindergarten and first grade. It was hypothesized that inhibitory control in preschool and caregiver involvement in kindergarten would mediate the associations between a history of maltreatment and foster placement and school adjustment in kindergarten and first grade. To meet the criteria for mediation, the hypothesized mediators needed to be significantly associated with the predictor and outcome variables (Baron & Kenny, 1986). This proved to be the case for inhibitory control, which was significantly negatively associated with a history of maltreatment and foster placement and significantly positively associated with the school adjustment measures. Also in accordance with predictions, caregiver involvement in school was significantly negatively associated with a history of maltreatment and foster placement and significantly positively associated with academic and social-emotional competence.

The associations among the covariates and the other measures were also examined. As detailed earlier, there were moderate correlations between some school characteristic variables and a history of maltreatment and foster placement; however, these variables were not significantly associated with the mediators or the outcomes. Thus, these variables were not used as covariates in the path analyses described next. General cognitive ability was strongly associated with academic competence and less strongly (but significantly) associated with social-emotional competence. Additionally, general cognitive ability and age at the baseline assessment were significantly positively associated with inhibitory control.

Path Analysis

Mplus was used to conduct the path analyses to determine whether the patterns of associations between a history of maltreatment and foster placement and the hypothesized mediators and outcomes remained significant when age and general cognitive ability were controlled. Additionally, the significance of mediated paths was tested. The results of the path analysis, trimmed to show only paths with a probability level of .10 or less, are shown in Figure 1. (The patterns of significance did not differ from those in the full, nontrimmed model, which was not a significantly better fit than the model presented here.) The trimmed model

Table 3

Correlations Among Hypothesized Mediators, Covariates, and Outcomes

	1	2	3	4	5	6	7	8	9	10
1. Maltreatment and foster care	—									
2. Academic competence	−.29**	—								
3. Social-emotional competence	−.29**	.33**	—							
4. Inhibitory control	−.34**	.35**	.33**	—						
5. Caregiver involvement	−.24**	.30**	.37**	.21*	—					
6. Percentage of students receiving free or reduced lunch	.17†	−.08	.07	−.12	−.06	—				
7. Average class size	.01	−.09	−.04	.06	−.17†	−.30**	—			
8. Percentage of teachers with a master's degree or higher	−.25**	.10	.05	−.01	−.06	−.04	.16†	—		
9. Child in private or magnet school	−.30**	.17†	−.05	.04	−.01	−.29**	−.01	.05	—	
10. General cognitive ability	−.37**	.50**	.17*	.39**	.12	−.18*	−.02	.20*	.11	—
11. Age at the baseline assessment	.13	−.06	.05	.38**	.14	−.07	.05	−.11	−.06	.04

$^\dagger p < .10$. $^*p < .05$. $^{**}p < .01$.

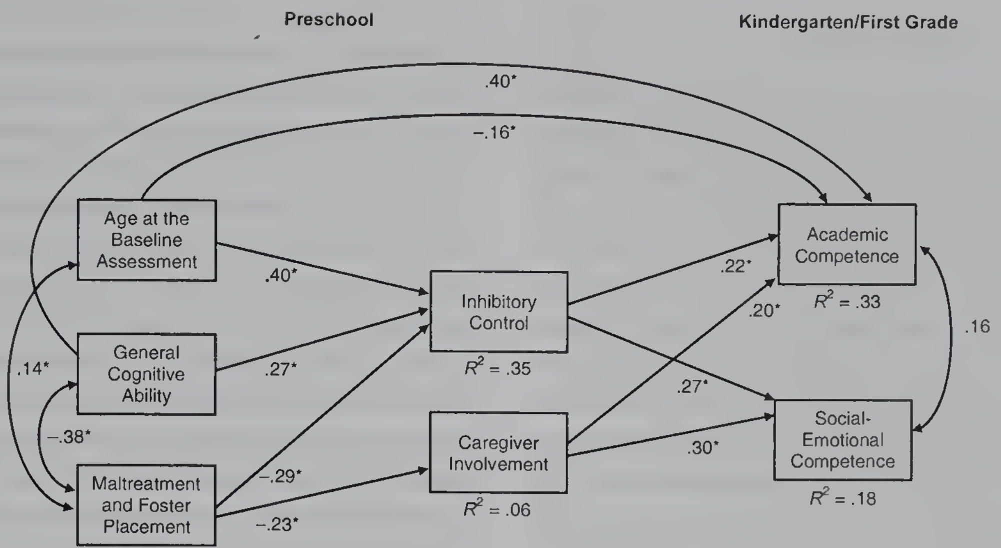


Figure 1. Path analysis of the associations between a history of maltreatment and foster placement, hypothesized mediators, covariates, and school adjustment.
$^*p < .05$.

showed a good fit, $\chi^2 = 8.84$, $p = .36$, comparative fit index (CFI) = .99, Tucker-Lewis index (TLI) = .99, root mean square error of approximation (RMSEA) = .000, standardized root mean square residual (SRMR) = .05. As was hypothesized, although a history of maltreatment and foster placement was significantly related to academic and social-emotional competence in the bivariate correlations, it was not in the full model. Direct paths between a history of maltreatment and foster placement and the school adjustment variables did not reach significance. Instead, as was predicted, the effects of a history of maltreatment and foster placement on academic and social-emotional competence were fully mediated through inhibitory control. The indirect path from a history of maltreatment and foster placement to academic competence through inhibitory control was significant ($t = -2.08$), as was the indirect path from a history of maltreatment and foster placement to social-emotional competence through inhibitory control ($t = -2.74$).

Additionally, there was a significant indirect path from a history of maltreatment and foster placement to social-emotional competence through caregiver involvement ($t = -2.10$). This finding supports the hypothesis that the effects of a history of maltreatment and foster placement on social-emotional competence are mediated through caregiver involvement. There was also a significant association between caregiver involvement and academic competence, although the indirect path from a history of maltreatment and foster placement to academic competence through caregiver involvement was not significant.

In the model, general cognitive ability was no longer directly associated with social-emotional competence. Instead, there was a significant indirect path from general cognitive ability to social-emotional competence through inhibitory control ($t = 2.38$), indicating full mediation. Age at the baseline assessment was significantly negatively associated with academic competence, suggesting that children who entered the study at younger ages demonstrated better academic competence.

To examine whether there were different patterns of associations within the FC and CC groups, a multiple groups model was run. The patterns of associations were similar across the two groups with two exceptions: General cognitive ability was not associated with inhibitory control in the CC group, and age at the baseline assessment was not associated with academic competence in the CC group. Also, because the FC and CC groups differed significantly on general cognitive ability (with a 1.78 point mean difference equal to .60 of a standard deviation, or about 9 IQ points), additional analyses were undertaken to determine if the group differences in school adjustment might be attributable to differences in general cognitive ability. Although the group means differed significantly, there was also considerable overlap between the two groups, allowing for exact matching of 40 FC children with 40 CC children on general cognitive ability scores. The path model described earlier was then rerun with only those 80 children. Although significance was in some cases reduced due to the smaller sample size, the patterns of associations were the same as those found in the path analysis that included the whole sample with general cognitive ability as a covariate.

Discussion

Maltreated foster children are at high risk for difficulties in school adjustment (Blome, 1997; Courtney & Dworksy, 2006; Geenen & Powers, 2006; Goerge et al., 1992; Smithgall et al., 2004). The current study demonstrates that these difficulties are apparent as early as school entry. Compared to their nonmaltreated peers, maltreated foster children showed lower academic and social-emotional competence in kindergarten and first grade. This finding was not surprising given that preschool-aged foster children show deficits in a number of domains such as language, psychosocial functioning, and neuropsychological functioning (Klee, Kronstadt, & Zlotnick, 1997; Pears & Fisher, 2005; Stahmer et al., 2005). However, this is the first study to focus on the functioning of maltreated foster children at school entry. The finding that maltreated foster children demonstrate deficits in the early school years underscores the significance of screening preschool-aged foster children for developmental delays. Additionally, the results from this study suggest that, in the absence of prior screening, foster children might need to be screened upon school entry for academic and social-emotional difficulties to enable timely intervention.

Preschool levels of inhibitory control were positively associated with academic and social-emotional competence in kindergarten and first grade. This finding provides further support for the hypothesis that inhibitory control abilities are important components of school readiness (McClelland et al., 2006) and is consistent with other studies linking inhibitory control with greater school readiness skills (e.g., Blair & Razza, 2007). As noted earlier, children who show greater inhibitory control are able to regulate prepotent behavioral and attentional responses. Thus, they are more likely to focus on academic activities. They are also more likely to control their behavioral responses to peers and to show more prosocial behavior.

Inhibitory control was negatively associated with a history of maltreatment and foster placement, suggesting that these experiences have deleterious effects on this ability. This is consistent with other studies showing that early parenting and family environment shape the development of inhibitory control in the toddler and preschool years (Kochanska et al., 2000; Valiente, Lemery-Chalfant et al., 2007). Specifically, this association supports findings from other populations of children who have experienced early adversity. As noted earlier, other studies have demonstrated negative effects of maltreatment and multiple foster placements on inhibitory control (Beers & DeBellis, 2002; Lewis et al., 2007). Researchers working with postinstitutionalized adoptees who experienced severe deprivation have shown that these children also display deficits in attentional processes in general (Kreppner, O'Connor, Rutter, & the English and Romanian Adoptees Study Team, 2001) and in inhibitory control specifically (Bruce, Tarullo, & Gunnar, 2009). Thus, this finding is consistent with previous results suggesting that early adversity has significant neurobiological consequences. By and large, these studies have not yet disentangled the effects of maltreatment and multiple caregiver disruptions. This remains an important topic for future research.

These neurobiological effects of early adversity might impact longer term outcomes given that the effects of a history of maltreatment and foster placement on school adjustment a number of years later were fully mediated by preschool inhibitory control. Although these results illustrate the potential long-term developmental consequences of maltreatment and foster placement, they also suggest a potentially powerful avenue for preventive intervention. Helping maltreated foster children to improve their inhibitory control skills prior to kindergarten might help to prevent academic and social difficulties and, in turn, place the children on a track of increasing school success. A recent school-based preventive intervention program aimed at reducing behavioral problems and promoting social-emotional competence was shown to improve inhibitory control abilities in elementary school children (Riggs, Blair, & Greenberg, 2003), suggesting that inhibitory control abilities are amenable to intervention efforts.

The results from this study also support the importance of increasing caregiver involvement in schooling as a means of promoting maltreated foster children's positive school adjustment. As with inhibitory control, caregiver involvement fully mediated the effects of a history of maltreatment and foster placement on social-emotional competence. Foster caregivers were less involved in children's schooling than were the biological parents of the nonmaltreated children. As noted earlier, this disparity might be due to a number of factors including the foster caregiver's lack of information about the child's school history and the relatively short duration of foster placements. Additionally, foster caregivers might not be aware of the processes that govern access to special services and might be unsure about whether they can authorize the children's participation in them (Choice et al., 2001). Thus, foster caregivers might benefit from additional education about the needs of foster children and ways to make contact with school personnel and to obtain services for the children.

Inhibitory control and caregiver involvement are not likely to be the only mediators in the association between a history of maltreatment and foster placement and school outcomes. Other factors such as developmental status, language abilities, and early literacy skills are likely to impact foster children's success in early elementary school. However, the findings of the current study suggest that interventions that include foci on inhibitory control skills and caregiver involvement in schooling, along with emphases on early literacy skills to increase academic competence and social-emotional skills to enhance social competence, might be particularly effective in improving maltreated foster children's early school adjustment. Head Start, one of the most widely recognized school readiness interventions, is focused on children's early literacy and social-emotional skills and caregiver involvement in school (Zigler & Styfco, 1995). This groundbreaking program has been found to increase literacy and social skills in low-income children (Abbott-Shim, Lambert, & McCarty, 2003; U.S. Department of Health and Human Services, 2001). More recently, several promising programs with foci on social-emotional skills and behavior regulation have shown positive impacts on children's social skills and behaviors (Domitrovich, Cortes, & Greenberg, 2007; Webster-Stratton, Reid, & Stoolmiller, 2008). Although these programs have demonstrated efficacy with general populations of children, none has been tested specifically with maltreated foster children, and none offers all components that seem to be important for these children. For example, although Head Start is focused on early literacy skills, social-emotional skills, and caregiver involvement, it lacks a focus on improving inhibitory control. Conversely, programs aimed at increasing behavior regulation are not focused on caregiver involvement. Thus, perhaps programs that emphasize inhibitory control and caregiver involvement as well as early literacy skills and social-emotional skills will be more efficacious in improving the school readiness of maltreated foster children (Pears, Fisher, & Bronz, 2007; Pears, Fisher, Heywood, & Bronz, 2007).

In the current study, it was surprising that school characteristics did not appear to be related to outcomes in correlational analyses. Other researchers have found such factors to be associated with school adjustment (Konstantopoulos, 2006). The maltreated foster children and their nonmaltreated peers were matched on SES, and the schools in the metropolitan area in which this study was conducted were all relatively well ranked. Thus, the schools might have been too homogeneous to detect the effects of school characteristics on outcomes. Future researchers should explore how school quality might help maltreated foster children to compensate for deficits in academic and social-emotional skills.

In this study, we identified two important mediators between a history of maltreatment and foster placement and early school adjustment. However, a few limitations and directions for future research should be noted. First, although the sample included a relatively large number of maltreated foster children and the effects of the mediators were moderate, the sample size might have limited the power to detect weaker associations among variables. The smaller sample size also precluded examining the effects of specific dimensions of maltreatment and foster care experiences (e.g., type of maltreatment and number of caregiver transitions) on the mediators and school outcomes. As noted earlier, maltreated foster children are unique because they have been maltreated and placed into foster care. Thus, future work should focus on disentangling the effects of specific maltreatment and foster care experiences. Additionally, as noted earlier, although inhibitory control and caregiver involvement appear to be important mediators of the association between a history of maltreatment and foster placement and later school adjustment, they are not likely to be the only ones. Future researchers should focus on identifying additional mediators that would provide further targets for preventive interventions. Finally, there were no gender differences in the outcomes or mediators in the current sample. Using a subsample of the children in the current study, Leve, Fisher, and DeGarmo (2007) found that gender moderated the association between foster care status and peer relations in the early grades. However, that study focused exclusively on measures of peer relations rather than on broader measures of social-emotional competence.

The findings from the current study demonstrate that maltreated foster children are markedly behind their peers in academic and social-emotional competence as early as kindergarten and first grade. In the absence of intervention, these deficits are likely to lead to maladaptive trajectories of academic achievement across the school years. There is cause for hope, however, in the identification of two possible targets for preventive intervention: inhibitory control and caregiver involvement. Interventions that target inhibitory control and caregiver involvement as well as early literacy and social-emotional skills might place the maltreated foster children on increasingly positive trajectories of academic and social success.

References

Abbott-Shim, M., Lambert, R., & McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a Head Start program and the program's wait list. *Journal of Education for Students Placed at Risk*, *8*, 91–214.

Anthonysamy, A., & Zimmer-Gembeck, M. J. (2007). Peer status and behaviors of maltreated children and their classmates in the early years of school. *Child Abuse and Neglect*, *31*, 371–991.

Barnard, W. M. (2004). Parent involvement in elementary school and educational attainment. *Child and Youth Services Review*, *26*, 39–62.

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. Toth (Eds.), *Child abuse, child development, and social policy* (Vol. 8, pp. 7–73). Norwood, NJ: Ablex.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Beers, S. R., & DeBellis, M. D. (2002). Neuropsychological function in children with maltreatment-related post-traumatic stress disorder. *American Journal of Psychiatry*, *159*, 483–486.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111–127.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *72*, 647–663.

Blome, W. W. (1997). What happens to foster kids: Educational experiences of a random sample of foster care youth and a matched group of non-foster care youth. *Child and Adolescent Social Work Journal*, *14*, 41–53.

Bruce, J., Tarullo, A. R., & Gunnar, M. R. (2009). Disinhibited social behavior among internationally adopted children. *Development and Psychopathology*, *21*, 157–171.

Burchinal, M., Roberts, J. E., Zeisel, S. A., Hennon, E. A., & Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in the early elementary school years. *Parenting: Science and Practice*, *6*, 79–113.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053.

Casey, B. J., Tottenham, N., & Fossella, J. (2002). Clinical, imaging, lesion, and genetic approaches toward a model of cognitive control. *Developmental Psychobiology*, *40*, 237–254.

Cassidy, J., & Asher, S. R. (1992). Loneliness and peer relations in young children. *Child Development*, *63*, 350–365.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, *98*, 489–507.

Child Welfare Information Gateway. (2007). *Foster care statistics*. Retrieved July 8, 2010, from http://www.childwelfare.gov/pubs/factsheets/foster.pdf

Choice, P., D'Andrade, A., Gunther, K., Downes, D., Schaldach, J., Csiszar, C., et al. (2001). *Education for foster children: Removing barriers to academic success*. Berkeley: University of California.

Christenson, S. L. (1999). Families and schools. In R. C. Pianta & M. J. Cox (Eds.), *The transition to kindergarten* (pp. 143–177). Baltimore: Paul H. Brookes.

Conduct Problems Prevention Research Group. (1995). *Technical reports for the construct development for the measures for Year 2 outcome analyses*. Unpublished technical report.

Courtney, M. E., & Dworksy, A. (2006). Early outcomes for young adults transitioning from out-of-home care in the USA. *Child and Family Social Work*, *11*, 209–219.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention*, *28*, 67–91.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Fisher, P. A., Gunnar, M. R., Chamberlain, P., & Reid, J. B. (2000). Preventive intervention for maltreated preschool children: Impact on children's behavior, neuroendocrine activity, and foster parent functioning. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 1356–1364.

Fothergill, K. E., Ensminger, M. E., Green, K. M., Crum, R. M., Robertson, J., & Juon, H. (2008). The impact of early school behavior and educational achievement on adult drug use disorders: A prospective study. *Drug and Alcohol Dependence*, *92*, 191–199.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*, 483–527.

Geenen, S., & Powers, L. E. (2006). Are we ignoring youths with disabilities in foster care? An examination of their school performance. *Social Work*, *51*, 233–241.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153.

Goerge, R. M., Van Voorhis, J., Grant, S., Casey, K., & Robinson, M. (1992). Special education experiences of foster children: An empirical study. *Child Welfare Journal*, *71*, 419–437.

Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology*, *45*, 3–19.

Greenberg, M., & Lengua, L. (1995). *Scale construction for the Seattle Personality Inventory*. Seattle: University of Washington.

Hooper, S. R., Roberts, J. E., Zeisel, S. A., & Poe, M. (2003). Core language predictors of behavioral functioning in early elementary school children: Concurrent and longitudinal findings. *Behavioral Disorders*, *29*, 10–24.

Howse, R. B., Calkins, S. D., Anastopoulos, A. D., Keane, S. P., & Shelton, T. L. (2003). Regulatory contributors to children's kindergarten achievement. *Early Education and Development*, *14*, 101–119.

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education*, *71*, 151–174.

Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school academic achievement. *Urban Education*, *40*, 237–269.

Klee, L., Kronstadt, D., & Zlotnick, C. (1997). Foster care's youngest: A preliminary report. *American Journal of Orthopsychiatry*, *67*, 290–299.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development*, *72*, 1091–1111.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*, 220–232.

Konstantopoulos, S. (2006). Trends of school effects on student achievement: Evidence from NLS:72, HSB:82, and NELS:92. *Teachers College Record*, *108*, 2250–2581.

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: Psychological Corporation.

Kreppner, J. M., O'Connor, T. G., Rutter, M., & the English and Romanian Adoptees Study Team (2001). Can inattention/overactivity be an institutional deprivation syndrome? *Journal of Abnormal Child Psychology*, *29*, 518–528.

Kurtz, P. D., Gaudin, J. M., Wodarski, J. S., & Howing, P. T. (1993). Maltreatment and the school-aged child: School performance consequences. *Child Abuse and Neglect*, *17*, 581–589.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373–1400.

Lamb-Parker, F., Piotrkowski, C. S., Baker, A. J. L., Kessler-Sklar, S., Clark, B., & Peay, L. (2001). Understanding barriers to parent involvement in Head Start: A research-community partnership. *Early Childhood Research Quarterly*, *16*, 35–51.

Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of Applied Developmental Psychology*, *28*, 40–55.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, *99*, 821–834.

Leslie, L. K., Landsverk, J., Ezzet-Lofstrom, R., Tschann, J. M., Slymen, D. J., & Garland, A. F. (2000). Children in foster care: Factors influencing outpatient mental health service use. *Child Abuse and Neglect*, *24*, 465–476.

Leve, L. D., Fisher, P. A., & DeGarmo, D. S. (2007). Peer relations at school entry: Sex differences in the outcomes of foster care. *Merrill-Palmer Quarterly*, *53*, 557–577.

Lewis, E., Dozier, M., Ackerman, J., & Sepulveda, S. (2007). The effect of caregiving instability on adopted children's inhibitory control abilities and oppositional behavior. *Developmental Psychology*, *43*, 1415–1427.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471–490.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Authors.

Nagin, D. S., & Tremblay, R. E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. *Archives of General Psychiatry*, *58*, 389–394.

Newman, J. P., Noel, A., Chen, R., & Mastopoulos, A. S. (1998). Temperament, selected moderating variables, and early reading achievement. *Journal of School Psychology*, *36*, 215–232.

NICHD Early Child Care Research Network. (2003). Do children's attention processes mediate the link between family predictors and school readiness? *Developmental Psychology*, *39*, 581–593.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*, *41*, 428–442.

Oregon Department of Education. (2007). *School and district report cards*. Retrieved from http://www.ode.state.or.us/search/page/?id=1786

Parrish, T., Delano, C., Dixon, D., Webster, D., Berrick, J. D., & Bolus, S. (2001). *Education of foster group home children, whose responsibility is it? Study of the educational placement of children residing in group homes*. Palo Alto, CA: American Institutes of Research.

Patterson, G. R., & Bank, L. (1986). Bootstrapping your way in the nomological thicket. *Behavior Assessment*, *8*, 49–73.

Pears, K. C., & Fisher, P. A. (2005). Developmental, cognitive, and neuropsychological functioning in preschool-aged foster children: Associations with prior maltreatment and placement history. *Journal of Developmental and Behavioral Pediatrics*, *26*, 112–122.

Pears, K. C., Fisher, P. A., & Bronz, K. D. (2007). An intervention to promote social emotional school readiness in foster children: Preliminary outcomes from a pilot study. *School Psychology Review*, *36*, 665–673.

Pears, K. C., Fisher, P. A., Heywood, C. V., & Bronz, K. D. (2007). Promoting school readiness in foster children. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on social learning in early childhood education* (pp. 173–198). Greenwich, CT: Information Age.

Peth-Pierce, R. (2000). *A good beginning: Sending America's children to school with the social and emotional competence they need to succeed.* Bethesda, MD: The Child and Mental Health Foundations and Agencies Network.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, *33*, 444–458.

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report*, *16*, 3–18.

Riggs, N. R., Blair, C. B., & Greenberg, M. T. (2003). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology*, *9*, 267–276.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. A. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*, 1394–1408.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology*, *33*, 906–916.

Smithgall, C., Gladden, R. M., Howard, E., Goerge, R., & Courtney, M. (2004). *Educational experiences of children in out-of-home care*. Chicago: Chapin Hall Center for Children at the University of Chicago.

Stahmer, A. C., Leslie, L. K., Hurlburt, M., Barth, R. P., Webb, M. B., Landsverk, J., et al. (2005). Developmental and behavioral needs and service use for young children in child welfare. *Pediatrics*, *116*, 891–900.

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion*, *7*, 77–88.

U.S. Department of Health and Human Services. (2001). *Head Start FACES: Longitudinal findings on program performance: Third progress report*. Washington, DC: Author.

Valiente, C., Lemery-Chalfant, K., & Castro, K. S. (2007). Children's effortful control and academic competence: Mediation through school liking. *Merrill-Palmer Quarterly*, *53*, 1–25.

Valiente, C., Lemery-Chalfant, K., & Reiser, M. (2007). Pathways to problem behaviors: Chaotic homes, parent and child effortful control, and parenting. *Social Development*, *16*, 249–267.

Veltman, M. W. M., & Browne, K. D. (2001). Three decades of child maltreatment research: Implications for the school years. *Trauma, Violence, and Abuse*, *2*, 215–239.

Walker, H. M., & McConnell, S. R. (1988). *The Walker-McConnell Scale of Social Competence and School Adjustment*. Austin, TX: Pro-Ed.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness. *Journal of Child Psychology and Psychiatry*, *49*, 471–488.

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scales of Intelligence–Revised*. San Antonio, TX: Psychological Corporation.

Zigler, E. F., & Styfco, S. (1995). *Head Start and beyond: A national plan for extended childhood intervention*. New Haven, CT: Yale University Press.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1565–1581

On the Progression and Stability of Adolescent Identity Formation: A Five-Wave Longitudinal Study in Early-to-Middle and Middle-to-Late Adolescence

Wim Meeus, Rens van de Schoot,
and Loes Keijsers
Utrecht University

Seth J. Schwartz
University of Miami

Susan Branje
Utrecht University

This study examined identity development in a 5-wave study of 923 early-to-middle and 390 middle-to-late adolescents thereby covering the ages of 12–20. Systematic evidence for identity progression was found: The number of diffusions, moratoriums, and searching moratoriums (a newly obtained status) decreased, whereas the representation of the high-commitment statuses (2 variants of a [fore]closed identity: ''early closure'' and ''closure,'' and achievement) increased. We also found support for the individual difference perspective: 63% of the adolescents remained in the same identity status across the 5 waves. Identity progression was characterized by 7 transitions: diffusion → moratorium, diffusion → early closure, moratorium → closure, moratorium → achievement, searching moratorium → closure, searching moratorium → achievement, and early closure → achievement.

Erikson (1968) theorized that one of the main tasks for adolescents is to develop a coherent sense of identity. Marcia's (1966) identity status model has been one of the most important, and widely studied and utilized, elaborations of Erikson's views on identity formation. Marcia distinguished four identity statuses based on the amount of exploration and commitment the adolescent is experiencing or has experienced. Identity diffusion (D) indicates that the adolescent has not yet made a commitment regarding a specific developmental task and may or may not have explored among different alternatives in that domain. Foreclosure (F) signifies that the adolescent has made a commitment without much prior exploration. In moratorium (M), the adolescent is in a state of active exploration and has not made significant commitments. Identity achievement (A) signifies that the adolescent has finished a period of active exploration and has made a commitment based on this exploration. Notably, Waterman (1982) has proposed that adolescents move from diffusion toward achievement as they progress through adolescence. This study was designed to test this developmental interpretation of the identity status model: Are identity statuses stable individual dispositions or do they change over time? We used a five-wave longitudinal data set to study identity formation from early to late adolescence (ages 12–20).

Identity Statuses: Individual Differences or Development?

In his original contribution, Marcia (1966) conceptualized the identity statuses in terms of individual differences: ''as individual styles of coping with the psychosocial task of forming an ego identity'' (p. 558). In this sense, the statuses represent different individual states or dispositions. Most identity researchers have adopted this perspective and view the statuses as stable individual dispositions. However, some writers have proposed that the statuses constitute a developmental sequence.

This study was supported by grants to Wim Meeus and Susan Branje from the Netherlands Organization for Scientific Research (NWO).

Correspondence concerning this article should be addressed to Wim Meeus, Research Centre on Adolescent Development, Utrecht University, PO Box 80.140, 3508 TC Utrecht, Netherlands. Electronic mail may be sent to w.meeus@uu.nl.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0018

Indeed, Waterman (1982) proposed a "developmental hypothesis" of the identity status model. This hypothesis involves two assumptions: First, the development of identity has a direction: development represents "changes in identity status that constitute progressive developmental shifts" (Waterman, 1982, p. 343). Such progressive development involves a movement away from diffusion and toward achievement. The second assumption is that progressive development also involves a specific pattern of transitions between identity statuses: from diffusion into foreclosure or moratorium, and from foreclosure and moratorium into achievement. Consequently, Waterman assumes that adolescents starting from diffusion move through foreclosure or moratorium and then into achievement. Most prominent among the identity transitions that Waterman proposes are D → M or D → F, F → M, and M → A.

Overviews of studies using identity status classifications have offered limited, but consistent, support for the first assumption within Waterman's (1982) developmental hypothesis. Meeus (1996) reported that in 17 of 25 studies reviewed, the prevalence of achievers was higher in the older age groups, and the prevalence of diffusions was higher in the younger age groups. In their reviews, both Van Hoof (1999) and Berzonsky and Adams (1999) found progressive developmental trends in 7 of 14 longitudinal studies, and across studies they found a higher prevalence of progressive than regressive shifts. Findings also revealed that, across studies, more than half of participants remained in the same identity status during the course of the study. In a recent meta-analysis of cross-sectional and longitudinal studies, Kroger (2007) found the prevalence of achievements to be about 1.5 times higher in emerging adults (ages 22–29) as compared to middle adolescents (aged 15), and the prevalence of diffusions to be about 1.3 times lower.

Support for Waterman's (1982) second assumption is very scarce. A limited number of longitudinal studies (Adams & Fitch, 1982; Dellas & Jernigan, 1987; Kroger, 1988; Meeus, Iedema, Helsen, & Vollebergh, 1999; Waterman, Geary, & Waterman, 1974; Waterman & Goldman, 1976; Waterman & Waterman, 1971) have tested whether transitions out of some statuses (e.g., diffusion) are more prevalent than transitions into those statuses. The available findings suggest that in cases where adolescents do change statuses, more adolescents move out of diffusion than into it (in about 30% of the studies), more into than out of achievement (in about 50% of the studies), more out of than into foreclosure (in about 22% of the studies), and more out of than into moratorium (in about 29% of the studies). Over-time stability of the identity statuses was 59%. These patterns suggest that identity status is more likely to remain stable than to change, and that when change does occur, this change tends to be reflected in movement out of diffusion, foreclosure and moratorium, and into achievement.

These findings are not inconsistent with Waterman's (1982) second assumption, but they also suggest that the assumption has yet to be subjected to rigorous empirical test. A full test requires that identity status changes are simultaneously tested in a transition table that incorporates all possible transitions between identity statuses; in a two-wave study with four identity statuses, this would require a test of a 4 × 4 transition matrix. The studies cited earlier only tested whether the number of movers into each status was different from the movers out of that status. As a result, systematic empirical tests of the sequential patterns of identity status transitions that Waterman hypothesized remain to be conducted, for instance, that the chances for the transition from diffusion to moratorium are greater than for the transition from diffusion to achievement.

The major aim of this study is therefore to provide a systematic account of identity status transitions—both progressive and regressive—over time. The prevalence of transitions will also provide information regarding the extent to which identity statuses represent stable individual dispositions or change over time. We now proceed to presenting our conceptualization of identity formation, in terms of the processes that underlie the identity statuses.

A Dimensional Approach: Commitment, In-Depth Exploration, and Reconsideration

Our approach focuses on the management of commitments and posits three dimensions as underlying the process of identity formation. *Commitment* refers to strong choices that adolescents have made with regard to various developmental domains, along with the self-confidence that they derive from these choices. *In-depth exploration* represents the ways in which adolescents maintain their present commitments. It refers to the extent to which adolescents actively explore the commitments that they have already made by reflecting on their choices, searching for information about these commitments, and talking with others about them. *Reconsideration of commitment* refers to the willingness to discard one's commitments and to search for new commitments. Reconsideration refers to the comparison of present commitments with possible alternative commitments when the present ones are no longer satisfactory.

Our model assumes that identity is formed in a process of continuous interplay between commitment, in-depth exploration, and reconsideration. Our model holds that individuals enter adolescence with a set of commitments of at least minimal strength in important ideological and interpersonal identity domains, and that adolescents do not begin the identity development process with a ''blank slate.'' The initial commitments build upon the way in which adolescents have resolved the earlier Eriksonian psychosocial crises in childhood and have developed the ego strengths of hope, will, purpose and competence (Erikson, 1968). Numerous studies have offered support for these assumptions. Markstrom, Sabino, Turner, and Berman (1997) and Markstrom and Marshall (2007) found clear links between previous Eriksonian ego strengths and identity achievement. Moreover, a number of studies have suggested that early adolescents can possess strong identity commitments (Adams & Jones, 1983; Archer, 1982; Meeus et al., 1999).

During adolescence, individuals manage their commitments in two ways: through in-depth exploration and through reconsideration. In-depth exploration is a process of continuous monitoring of present commitments and serves the function to make them more conscious and to maintain them. Reconsideration is the process of comparing present commitments to alternative ones and deciding whether they need to be changed. Our model therefore focuses on the dynamic between certainty (exploration in depth) and uncertainty (reconsideration).

So, our model differs from Marcia's (1966) model in two respects. First, it differentiates Marcia's concept of exploration into in-depth exploration and reconsideration, which serve to maintain and change commitments, respectively. Second, our model has a stronger process orientation than Marcia's model. Marcia views commitments as the outcome of the process of exploration: After exploring various alternative commitments, adolescents choose one or more to which they will adhere. In contrast, our model assumes, as suggested by Grotevant (1987, p. 214), that commitments are formed and revised in an iterative process of choosing commitments and reconsidering them. In addition, our model assumes that adolescents regularly reflect upon their present commitments. In sum, our conceptualization of the process of identity formation implies a twofold management of present commitments. This conceptualization of in-depth exploration and reconsideration resembles the distinction between exploration in depth and exploration in breadth that was originally suggested by Grotevant (1987) and that has been applied by Luyckx, Goossens, and Soenens (2006) in their dual-cycle model of identity formation.

By including commitment, exploration in depth, and reconsideration in our model, we sought to capture Erikson's (1968) dynamic of *identity versus identity diffusion*. Commitment and in-depth exploration on the one hand, and reconsideration on the other hand, are conceptualized as the two opposing forces within this dynamic: whereas commitment and in-depth exploration imply attempts to develop and maintain a sense of self (i.e., identity coherence or synthesis), reconsideration represents questioning and rethinking this sense of self (identity confusion). To measure this three-dimensional model of identity formation, we developed the Utrecht-Management of Identity Commitments Scale (U-MICS; Crocetti, Rubini, and Meeus, 2008) as an extension of the earlier Utrecht-Groningen Identity Development Scale.

As was the case with Marcia's (1966) original dimensions of exploration and commitment, our three-dimensional model can be used to assign participants to identity status categories. For example, using cluster-analytic procedures in a cross-sectional study among 1952 Dutch early and middle adolescents, Crocetti, Rubini, Luyckx, and Meeus (2008) extracted five statuses from continuous measures of commitment, in-depth exploration, and reconsideration. Four of these statuses very closely resembled Marcia's four statuses. Achievement was represented as a combination of high commitment, high in-depth exploration, and very low reconsideration. Moratorium was represented by a combination of relatively low commitment, moderate in-depth exploration, and relatively high reconsideration; foreclosure as high commitment, relatively low in-depth exploration, and very low reconsideration; and diffusion as very low commitment, very low in-depth exploration, and very low reconsideration. In addition to these four statuses, a fifth status also emerged—a combination of high commitment, high in-depth exploration, and very high reconsideration. Crocetti, Rubini, Luyckx, et al. (2008) labeled this status as *searching moratorium* (SM). Adolescents in this status have strong commitments and explore them intensively, but they are also very active in considering alternative commitments. Crocetti, Rubini, Luyckx, et al. found that SMs were empirically distinguishable from "classical" moratoriums in terms of psychosocial functioning—compared to those in classical moratorium, individuals classified into SM were characterized by lower levels of depression, anxiety, and aggression, as well as by more favorable relationships with parents. These findings underscore the differences between the two moratorium statuses and suggest that SMs seek alternative commitments while already possessing strong commitments, whereas classical moratoriums do so with weak or no current commitments.

Within this context, it is important to note that the foreclosed status, as defined by Crocetti, Rubini, Luyckx, et al. (2008), may carry a different psychological meaning depending on the developmental pathways through which adolescents arrive at this status. As a result, in this study, we differentiate foreclosure into two subtypes—"early closures" and "closures." When adolescents begin in the foreclosed status and remain there over time, they can be considered to be "early closures," given that they have strong commitments that were established early on, have not tried to consider alternative commitments, and have not engaged in in-depth exploration of their present commitments. Adolescents, however, also can move from moratorium to this status of high commitment, low in-depth exploration, and low reconsideration. In this case, they have considered alternative commitments, are not engaged in in-depth exploration of present commitments, and should be labeled simply as "closures." Among adolescents with high commitments and low levels of in-depth exploration and reconsideration, we expected that the longitudinal clustering procedures used in the present study would be able to distinguish between closures and early closures.

The differentiation between closures and early closures is intended to highlight the similar profile, but different developmental roots, of these two subtypes of foreclosure. As a result, when we refer to both types of closure, we use the label early closure/closure (ECC). Separate labels—EC for early closure and C for closure—are used to refer to the distinct statuses.

Taken together the findings reported by Crocetti, Rubini, Luyckx, et al. (2008) suggest that our three-dimensional model yields identity statuses that are conceptually quite similar to those proposed by Marcia. Therefore, as is the case with Marcia's statuses, the statuses found with the three-dimensional model can be ordered on an identity status continuum. Diffusion and achievement represent the least and most mature endpoints of the continuum, respectively, with moratorium, SM, and ECC representing intermediate statuses: D–M–SM–ECC–A. Therefore, the identity status continuum generated by the three-dimensional model offers the potential to study change and stability of identity status.

The Present Study: Aims and Hypotheses

The primary goal of this study was to evaluate the extent to which identity statuses represent stable individual dispositions versus states into and out of which individuals move over time during adolescence. Both assumptions of Waterman's (1982) "developmental hypothesis" were evaluated here. Support for the first assumption would take the form of decreases in diffusion, and increases in achievement, over time. Support for the second assumption would take the form of the progressive transitions that Waterman proposed: D → M, D → F, F → M, and M → A. In the terminology used within our identity model, these transitions would be labeled as D → M, D → ECC, ECC → M, and M → A. Based upon the prior literature, we expected to find support for the first assumption. Literature examining the second assumption is fairly scarce, so we treated this as an exploratory research question. These issues were examined using data from a five-wave study, including an early-to-middle adolescent cohort and a middle-to-late adolescent cohort, thereby covering the ages from 12 to 20.

We also examined gender differences in identity statuses and identity transitions. In a review of identity status studies between 1966 and 1995 Kroger (1997) discussed gender differences in overall, interpersonal, and ideological identity. As is common in the identity status literature, she defined overall identity as ego strength and ego synthesis that individuals derive from commitments in a combination of life domains, and interpersonal and ideological identity as ego strength and synthesis that individuals derive from interpersonal and educational or work or political commitments, respectively. Kroger reported no gender differences in overall identity, but she did find that females were more often in achievement in interpersonal identity than males, and that in high school samples males seemed to move into the direction of achievement later than females (Kroger, 1997, pp. 752 and 754, respectively). Studies conducted since 1995 have replicated these findings with regard to interpersonal identity domains such as friendships (Lewis, 2003) and also have found a higher prevalence of females in achievement, along with a higher number of males in diffusion both in overall identity (Guerra & Braungart-Rieker, 1999) and in ideological identity domains (Schwartz & Montgomery, 2002). This pattern of findings suggests that since the 1960s, gender differences appear more often in interpersonal identity than in overall and ideological identity, and more often in high school samples than in college or university samples. Moreover, the more recent findings suggest that gender differences also may be more likely to appear in overall and ideological identity from the late 1990s on.

In this study, we used a Dutch sample where a majority of participants were in high school, and we used and a measure of overall identity that combines interpersonal and ideological domains. In the Netherlands, females may have stronger educational commitment because they have tended to perform better in school than males since the late 1990s (Statistics Netherlands, 2008b, 2008c). Additionally, Dutch females have been found to have stronger interpersonal commitments than their male counterparts (Meeus & Deković, 1995). Given the age and the nationality of our participants and our use of a combination of interpersonal and ideological domains to tap overall identity, we therefore would expect females, compared to males, to be more strongly represented in achievement and less so in diffusion.

Method

Participants

Data for this study were collected as part of an ongoing research project on COnflict And Management Of RElationships (CONAMORE; see Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2009), with a 1-year interval between each of the five available waves. The longitudinal sample consisted of 1,313 participants divided into an early-to-middle adolescent cohort ($n = 923$; 70.3%), who were 12.4 years of age ($SD = 0.59$) on average at baseline, and a middle-to-late adolescent cohort ($n = 390$; 29.7%) with an average age of 16.7 years ($SD = 0.80$) during the first wave of measurement. Because both age groups were assessed during five measurement waves, a total age range from 12 to 20 years was available. The early-to-middle adolescent cohort consisted of 468 boys (50.7%) and 455 girls (49.3%), and the middle-to-late adolescent cohort consisted of 169 boys (43.3%) and 221 girls (56.7%). In both the younger and older cohorts, the vast majority of adolescents (85.1% and 84.3%, respectively) indicated that they were living with both their parents. The remainder of adolescents lived with their mother (7.9% and 7.2% in the younger and older cohort, respectively) or elsewhere (e.g., with their father, with one biological parent and one stepparent, or with other family members). The composition of the two cohorts did not significantly differ with regard to ethnicity. In the younger cohort, 83.4% identified themselves as Dutch, and 16.6% indicated that they belonged to ethnic minorities (e.g., Surinamese, Antillean, Moroccan, Turkish). In the older cohort, 87.4% of participants were Dutch, and 12.6% were ethnic minorities. In the year when the current study was initiated (2001), 21% of all Dutch early-to-middle adolescents, and 22% of the Dutch middle-to-late adolescents, belonged to ethnic minority groups (Statistics Netherlands, 2008a). Thus, ethnic minorities were slightly underrepresented in our sample. With regard to education, all participants initially were in junior high and high schools. Given the Dutch educational system, most participants switched schools at least once during the study. Specifically, participants in the younger cohort switched from junior high school to high school, whereas most of the participants in the older cohort switched from high school to college or university. Because of the sample recruitment procedure, 100% of our middle-to-late adolescents were in high school or college, whereas national demographic statistics (Statistics Netherlands, 2008a, 2008b, 2008c) reveal that 96% of the Dutch middle-to-late adolescents was in some form of education during the time period covered by the current study (i.e., 2001–2005).

Sample attrition was 1.2% across waves: In Waves 1, 2, 3, 4, and 5, the numbers of participants were 1,313, 1,313, 1,293, 1,292 and 1,275, respectively. Missing values were estimated in SPSS (SPSS Inc., Chicago, IL), using the expectation maximization procedure. Little's missing completely at random test produced a normed χ^2 (χ^2/df) of 1.55, which, according to Bollen (1989), indicates that the data were likely missing at random, and that it is safe to impute missing values.

Procedure

Participating adolescents were recruited from various high schools in the province of Utrecht, Netherlands. Participants and their parents received an invitation letter describing the research project and goals and inviting them to participate. More than 99% of the families who were approached signed the informed consent form. During regular annual assessments, participating adolescents completed questionnaires at school or at home. Confidentiality of responses was guaranteed. Adolescents received €10 (approximately US$13) for each wave in which they provided data.

Measures

Identity was assessed using the U-MICS (Crocetti, Rubini, & Meeus, 2008). The U-MICS consists of 13 five-point Likert-scale items (1 = *completely untrue* to 5 = *completely true*), measuring identity in three dimensions: commitment, in-depth exploration, and reconsideration of commitment. Within the ideological and interpersonal domains, the U-MICS consists of five items measuring commitment, five items measuring in-depth exploration of present commitments, and five items measuring reconsideration of commitment. Sample items for the ideological and interpersonal domain include, respectively, ''My education/best friend makes me feel confident about myself'' (commitment), ''I often think about my education/best friend'' (in-depth exploration), and ''In fact, I'm looking for a different education/a new best friend'' (reconsideration of commitment). Although the U-MICS measures identity in different domains, the instrument can be used to assess overall identity. Crocetti, Rubini, Luyckx, et al. (2008) included both ideological (education) and interpersonal (best friend) domains and demonstrated the internal factorial validity of the three-dimensional model across domains. In this study, Cronbach's alphas for commitment, in-depth exploration and reconsideration of commitment ranged in both cohorts from .91 to .93, .88 to .89, and .84 to .94 across waves, respectively.

Analytic Strategy

To address our research questions, we utilized two applications of the general latent class model: latent class analysis (LCA) and latent transition analysis (LTA). Because we wanted to compare the prevalence of the various identity statuses between the early-to-middle and middle-to-late cohorts, we assumed measurement invariance across cohorts; that is, we restricted the profiles of the five identity classes on the three identity dimensions to be the same across cohorts. LCA is a person-centered analytic strategy that is a confirmatory version of cluster analysis. LCA groups individuals into classes based on empirically distinct patterns of scores on the variables (in this case the three identity dimensions) used to create the classes. LCA of continuous variables is sometimes referred to as latent profile analysis. For simplicity, we use the term LCA here. Like confirmatory factor analysis, LCA generates both measurement and structural parameters (Nylund, Asparouhov, & Muthén, 2007). The continuous scores for each of the identity variables within each class represent the measurement parameters, whereas the structural parameters refer to the class membership probabilities assigned to groups of individuals. Unlike cluster analysis, LCA offers fit statistics and significance tests to determine number of classes, it assigns class membership based on class probabilities, thereby taking uncertainty of membership, or error, into account. LCA has been found to be superior to cluster analysis in several Monte Carlo studies (Reinke, Herman, Petras, & Ialongo, 2008). In this study, we applied LCA to test whether the five hypothesized identity statuses would emerge in each of the five waves of measurement.

Latent transition analysis represents a longitudinal extension of LCA (for a recent overview of LTA, see Kaplan, 2008). LTA calculates patterns of stability and change over time in the form of movement or transitions between classes (identity statuses in this case). Like LCA, LTA models use class-specific parameters (the continuous scores for each of the identity variables within each class) as measurement parameters, and class probabilities as structural parameters to estimate the number of participants in each of the classes. To model change over time, LTA adds a second set of structural parameters, latent transition probabilities, to the latent class model. In a two-wave LTA, for example, transition probabilities refer to the probability of moving into class Y in Wave 2 conditional on having been in class X in Wave 1. These transition probabilities range between 0 and 1. In sum, then, LTA offers two types of structural parameters: (a) varying numbers of participants in class across waves, indicating increase or decrease in class size over time, and (b) transitions of individuals between classes that carry these changes of class size. Therefore, LTA is appropriate for evaluating both assumptions of Waterman's (1982) developmental hypothesis, the hypothesized decrease of diffusion and increase of achievement, and the hypothesized identity transitions that carry this increase or decrease over time.

Latent transition analysis results can be converted into contingency tables summarizing the prevalence of classes (identity statuses) across waves. We use Bayesian model selection using (in)equality constraints between the parameters of interest to evaluate these contingency tables. For a more detailed description of this method, readers are referred to Laudy and Hoijtink (2007). Using constraints may express prior information explicitly. This way we can evaluate the likelihood of certain patterns of increases and decreases in identity status membership. Moreover, expected differences in prevalence of identity statuses between early-to-middle and middle-to-late adolescents and males and females can be evaluated.

The results of the Bayesian model selection are expressed in terms of Bayes factors (BFs), representing the amount of evidence in favor of the model at hand compared to another model, and posterior model probabilities (PMPs), representing the probability that the model at hand is the best among a set of finite models after observing the data. PMPs of a model are computed by dividing its BF by the sum of all BFs.

Results

We present our results in four steps. First we apply cross-sectional LCA to explore the number of classes (identity statuses) within each of the five waves. Second, we select the best fitting five-wave LTA model in a number of successive steps. Because we wanted to compare the prevalence of the various identity statuses between the early-to-middle and middle-to-late cohorts, we assumed measurement invariance across cohorts by restricting the profiles of the five identity classes on the three identity dimensions to be the same across cohorts. Third, we apply Bayesian evaluations of the contingency tables generated by the final LTA model. The Bayesian evaluations are intended to address four research questions: (a) Is there a differential increase and decrease of identity statuses across time? (b) Are there differences in the prevalence of statuses across time between the early-to-middle and middle-to-late adolescents? (c) Likewise, are there differences in the prevalence between and males and females? (d) Is there a differential increase and decrease of identity statuses over time between males and females? Fourth, we globally describe the sequence of identity statuses in five-wave identity status trajectories (e.g., moratorium → moratorium → achievement → achievement → achievement in Waves 1 to 5, respectively).

Cross-Sectional Latent Class Analysis

For each of the five waves, we estimated a set of cross-sectional LCAs on the entire sample, including all three identity dimensions simultaneously. Analyses were performed using Mplus (Muthén & Muthén, 2007). We used four criteria to determine the number of latent classes (Nagin, 2005). First, a solution with k classes should result in improvement of model fit compared to a solution with $k - 1$ classes, indicated by a decrease of the Bayesian information criterion (BIC). Second, adding an additional class should lead to a *significant* increase of fit, as indicated by the bootstrap Lo–Mendel–Rubin likelihood ratio test (BLRT; Nylund et al., 2007). Third, entropy—a standardized measure of classification of individuals into classes, based upon the posterior class probabilities—of the final class solution should be acceptable. Entropy values range from 0 to 1, with values of .70 or higher indicating good classification accuracy (Reinecke, 2006). Fourth, we evaluated the content of the classes in the various solutions. If an additional class in a solution with k classes was found to be a slight variation of a class already found in a solution with $k - 1$ classes, we would choose the most parsimonious solution.

As expected, we found the five-class solution to be superior to the one- to four-class solutions on both fit indices across waves. BIC of the five-class solutions was at least 56.21 lower than that of one-, two-, three-, or four-class solutions and only in Wave 5 did the BLRT indicate that the five-class solution did not fit significantly better than the four-class solution ($p = .14$). Entropy (E) for the five-class solution ranged between .74 and .81, indicating good classification accuracy. Adding a sixth class did not provide additional unique information, given that the sixth class was small ($10 > n > 27$ in each wave) and appeared to represent a variation of one of the other classes. Therefore, we decided to use a five-class model in the LTA's.

Five-Wave Latent Transition Analysis

As part of the LTA, we assumed measurement invariance in the five-class LCA solutions across measurement waves. That is, we restricted the profiles of the five identity classes on the three identity dimensions to be equivalent across five waves. We also restricted the variances of the three identity dimensions to be equivalent across classes across waves. Assuming measurement invariance ensures that the profiles of the classes are the same across waves and allows for a straightforward interpretation of transition probabilities (see Nylund, Muthén, Nishina, Bellmore, & Graham, 2006). Figure 1 displays

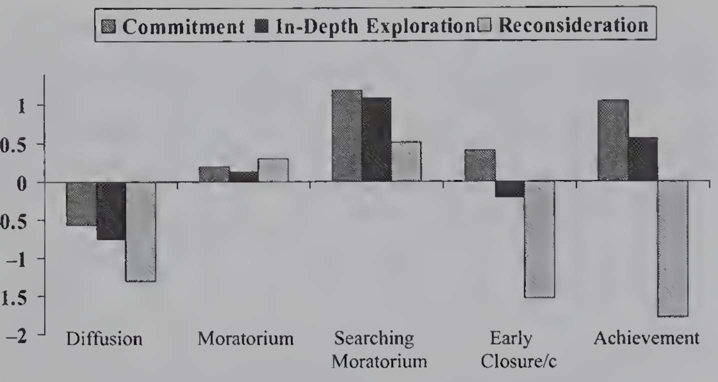


Figure 1. Profiles of the identity statuses on the three identity dimensions across waves.
Note. For reasons of presentation, the scores of the identity dimensions of the statuses were centered at a common scale point (2.75). Early closure/c = early closure/closure.

the profiles of the statuses. Four of the classes (achievement, moratorium, ECC, and diffusion) strongly resemble Marcia's (1966) original statuses. The fifth status, SM, combines very strong commitment with high levels of in-depth exploration and very high levels of reconsideration.

We developed the final LTA model in two steps. We will describe these steps and then present the results of the final model. In both steps, we selected the model with the lowest BIC value. The BLRT is not available for LTA models.

LTA Step 1: Nonstationary versus stationary transition probabilities. In the first modeling step, we compared a model with nonstationary transition probabilities between adjacent waves to a model with stationary transition probabilities. A model with nonstationary transition probabilities assumes that the likelihoods of transitions between classes are different between waves. In contrast, a model with stationary transition probabilities assumes that the probabilities are equal across waves. Results indicated no significant differences in the transition probabilities across time. The BIC for the LTA model with stationary transition probabilities (32,895.71) was lower than BIC of the model with nonstationary transition probabilities (33,112.67). This suggests that adolescents make transitions between identity statuses at the same pace across the four transitions points. As a result, there seems to be a very regular pattern of identity development. Entropy of the stationary model was very good, .85.

LTA Step 2: Are there age and gender differences in identity status transitions? We added covariates to the model with stationary transition probabilities to describe heterogeneity in transitions between statuses. In the first model, we included cohort as covariate to test whether transitions into and out of identity statuses were different between the early-to-middle and middle-to-late adolescents. The second model tested whether transitions were different for males and females. The first model comparison indicated no significant differences in the transition probabilities between the cohorts. The BIC for the LTA without cohort as covariate (32,895.71) was lower than BIC of the model with covariate (32,901.75). The second model comparison indicated significant gender differences in transition probabilities. The BIC for the LTA with gender (32,894.71) was lower than BIC of the model without gender (32,895.71), indicating that the model with gender was 2.72 times more likely than the model without gender (Nagin, 1999). So, rate of change into and out of identity statuses was not different for early-to-middle and middle-to-late adolescents, but was for males and females. Next, we present follow-up Bayesian analyses to clarify the gender differences.

Increase and decrease of identity statuses over time. Table 1 displays the cell sizes for each the five identity statuses for Waves 1, 2, 3, 4, and 5 based on the final LTA model. Findings for the whole sample are in the upper panel of the table. This table indicates a systematic decrease in D, M and SM over time, along with a systematic increase in ECC and A. The table also suggests that ECC is the most prevalent status: Between 50.6% and 55.2% of the sample was classified into that status across waves. A majority of the adolescents had relatively strong commitments along with relatively low levels of in-depth exploration and very low levels of reconsideration. The systematic pattern of increases and decreases in status membership is also found across both cohort and gender (Panels 2 and 3, and 4 and 5 of Table 1, respectively).

We applied Bayesian model selection (Laudy & Hoijtink, 2007) to the upper panel of Table 1 to test which of three alternative models of increase and decrease of identity status best fit the data. Model 1 assumed no increase or decrease of identity statuses across five waves, whereas Model 2 assumed a decrease of D, M, and SM and an increase of ECC and A. In Model 3, the unconstrained model, the distribution of statuses over time was allowed to vary freely. The results are in Table 2. First, Models 1 and 2 were compared to the unconstrained (Model 3). The BFs for Models 1 and 2 imply that after observing the data, these models are approximately 270 and 7,500 times as likely, respectively, as the unconstrained (Model 3). The second comparison revealed that Model 2 is 27.64 times as

Table 1
Size of Identity Status Classes for the Whole Sample, Early-to-Middle and Middle-to-Late Adolescents, and Males and Females

	Identity status									
	Diffusion		Moratorium		Searching moratorium		Early closure/closure		Achievement	
Wave	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Total sample (*N* = 1313)										
1	117	8.9	215	16.4	78	5.9	665	50.6	238	18.1
2	106	8.1	218	16.6	67	5.1	676	51.5	246	18.7
3	97	7.4	211	16.1	60	4.6	684	52.1	261	19.9
4	90	6.9	186	14.2	41	3.1	713	54.3	283	21.6
5	82	6.2	183	13.9	23	1.8	725	55.2	300	22.8
Early-to-middle adolescence (*N* = 923)										
1	86	9.3	159	17.2	70	7.6	455	49.3	153	16.6
2	76	8.2	162	17.6	64	6.9	463	50.2	158	17.1
3	68	7.4	146	15.8	57	6.2	481	52.1	171	18.5
4	66	7.2	127	13.8	36	3.9	505	54.7	189	20.5
5	61	6.6	131	14.2	23	2.5	510	55.3	198	21.5
Middle-to-late adolescence (*N* = 390)										
1	31	7.9	56	14.4	8	2.1	210	53.8	85	21.8
2	30	7.7	56	14.4	3	0.8	213	54.6	88	22.6
3	29	7.4	65	16.7	3	0.8	203	52.1	90	23.1
4	24	6.2	59	15.1	5	1.3	208	53.3	94	24.1
5	21	5.5	52	13.3	0	0.0	215	55.1	102	26.2
Males (*N* = 637)										
1	69	10.8	136	21.4	55	8.6	292	45.8	85	13.3
2	61	9.6	141	22.1	45	7.1	300	47.1	90	14.1
3	59	9.3	142	22.3	41	6.4	298	46.8	97	15.2
4	57	8.9	124	19.5	29	4.6	311	48.8	116	18.2
5	51	8.0	115	18.1	17	2.7	328	51.5	126	19.8
Females (*N* = 676)										
1	48	7.1	79	11.7	23	3.4	373	55.2	153	22.6
2	45	6.7	77	11.4	22	3.3	376	55.6	156	23.1
3	38	5.6	69	10.2	19	2.8	386	57.1	164	24.3
4	33	4.9	62	9.2	12	1.8	402	59.5	167	24.7
5	31	4.6	68	10.1	6	0.9	397	58.7	174	25.7

Note. Findings based on the final stationary 1-year interval model.

likely as Model 1. PMPs of Models 1, 2, and 3 are .03, .97, and < .001, respectively. Note that we assume that before observing the data each model is equally likely. In sum, Model 2, assuming decreases in D, M, and SM and increase in ECC and A, was by far the best fitting model. This model appears to support the first assumption of Waterman's (1982) developmental hypothesis.

Transitions between identity statuses. Transition probabilities of identity status change across 1-year intervals, as found in the final stationary model, are in Table 3. The transition probabilities of identity change between Waves 1 and 5 are displayed on the right-hand side of the table. The 4-year probabilities were calculated using the contingency tables of Waves 1 and 5 as generated by the final LTA model. We added these longer term transition probabilities as a way of elucidating change in identity status across a longer period of time. As expected given the consistency of identity status transition probabilities across time, the transitions during 1-year intervals strongly parallel the transitions during 4-year intervals. Transitions with a relatively high frequency during 1-year intervals were also highly likely during the 4-year interval. Not surprisingly, stability of identity statuses was greater during 1-year intervals than during the 4-year interval, and transitions between identity statuses were more likely to have occurred across 4 years than across 1 year.

Table 2
Bayesian Model Selection: Comparison of Various Sets of Models

	Model comparisons		
Models	BF		PMP
RQ1: Increase and decrease of identity statuses?			
M1. No increase or decrease of D, M, SM, ECC, and A	271.16	1	.03
M2. Decrease of D, M, and SM and increase of ECC and A	7,497.52	27.64	.97
M3. Unconstrained	1[a]		< .001
RQ2: Prevalence of identity statuses different between cohorts?			
M1. No difference in prevalence	0.004	1	< .001
M2. Systematic difference in prevalence	211.50	52,875	.99
M3. Unconstrained	1		< .01
RQ3: Prevalence of identity statuses different between males and females?			
M1. No difference in prevalence	< 0.01	1	< .001
M2. Systematic difference in prevalence	510.73	51,073	.99
M3. Unconstrained	1		< .001
RQ4: More decrease of D, M and SM and increase of ECC and A in males then females?			
M1. No gender-specific increase or decrease	0.004	1	< .001
M2. Gender-specific increase or decrease	5.29	1,322.5	.84
M3. Unconstrained	1		.16

Note. BF = Bayes factor; ECC = early closure/closure; PMP = posterior model probability.
[a]Models with BF = 1 are reference category.

Table 3
Transition Probabilities of Identity Status Change During 1-Year Intervals (n + 1) *and 4-Year Intervals* (n + 4) *Across Five Waves*

	Identity status in year, *n* + 1[a]					Identity status in year, *n* + 4				
Identity status in year, *n*	D	M	SM	ECC	A	D	M	SM	ECC	A
Diffusion (D)	.70	.09	.00	.19	.02	.39	.11	.00	.45	.05
Moratorium (M)	.03	.71	.04	.13	.09	.05	.39	.04	.30	.22
Searching moratorium (SM)	.00	.22	.50	.08	.19	.03	.26	.18	.22	.32
Early closure/closure (ECC)	.02	.03	.00	.90	.05	.03	.06	.00	.80	.11
Achievement (A)	.01	.05	.02	.11	.81	.01	.11	.00	.26	.62

Note. Findings of the final stationary model.
[a]For a stationary model, all transitions probabilities are the same across waves.

Seven specific findings warrant mention here. First, 1-year stability is always more likely than change in identity status. This is also true for 4-year stability of M, ECC, and A, and with two exceptions for the 4-year stability of D and SM, as compared to 4-year identity status change. Notably, 1- and 4-year stability probabilities for ECC and A are very substantial. Second, very few adolescents shifted from M, ECC, and A into D or SM; during 1-year intervals 4% or fewer of the adolescents made this transition, and during the 4-year interval, 5% or fewer did so. Third, transitions into moratorium are limited: during 1-year intervals, 9% of diffusions and 23% of SMs moved into moratorium, whereas during the 4-year interval 11% of diffusions and achievers, and 26% of SMs, moved into moratorium. Fourth, the percentage of transitions into ECC is substantial: between 8% and 19% of the adolescents in D, M, SM, and A moved into ECC during 1 year, and between 22% and 45% did so over the 4 years of the study. The likelihood of moving from D to ECC was very substantial, 19% and 45% during 1 and 4 years, respectively. Given the distinction we have made between early closure and closure in the introduction, the transition out of diffusion should be labeled as D → EC, given that adolescents who make this transition have never reconsidered identity alternatives or explored present commitments in depth. Similarly, remaining in the early closure status should be labeled as EC → EC. However, adolescents who move from SM, M, or A into ECC have considered identity alternatives or have explored present commitments in depth. Therefore, these transitions should be labeled as SM → C, M → C, and A → C. These refer to adolescents who once maintained high levels of in-depth exploration (in the case of regression from achievement to closure) or reconsideration (in the case of movement out of either of the moratorium statuses) and now report low levels of both in-depth exploration and reconsideration. Fifth, few adolescents move from D to A: 2% and 5% during 1 and 4 years, respectively. Sixth, transitions from EC, M, and especially SM into A are quite prevalent: from 5% to 19% and 11% to 32% over 1 and 4 years, respectively. This suggests that the likelihood of a diffused adolescent reaching achievement during adolescence is very low. Seventh, most of the transitions are progressive: from D, M, and SM to the direction of ECC and A. But there are also "regressive" transitions: notably from A to M and C, and from SM to M: 5% and 11%, 11% and 26%, and 22% and 26% during 1 and 4 years, respectively.

In addition, to check whether the stability of the findings is affected by the difference in sample size between cohorts, we estimated a replication of our final stationary model, controlling for sample size differences between cohorts by weighting the cohorts equally in the model. This was done by assigning the weight of 1 to each of the 923 cases of the younger cohort and the weight of 2,366 to each of the 390 cases in the older cohort. The replication confirmed all the earlier reported findings and suggests that they were not affected by difference in sample size between cohorts.

Identity status trajectories. Inspection of the five-wave identity status trajectories revealed two general patterns. First, 822 adolescents were in the same identity status in Waves 1 and 5. The vast majority of the 822 participants (93.7%, or 63% of the total sample) stayed in the same status in all waves. Second, 491 adolescents were in different statuses in Waves 1 and 5: 78.2% of them made only one status transition, 20.4% made two transitions, and 1.4% made three or more transitions during the five waves of the study. So the majority of identity status changers made only one transition. We also found that 11% of the change trajectories were two-transition trajectories in which adolescents passed through SM or M as transitory identity statuses.

These findings partially support the second assumption of Waterman's (1982) developmental hypothesis. We indeed found three of the four progressive identity status trajectories that Waterman hypothesized, notably D → M, D → EC, and M → A. We did not find general support for Waterman's assumption that adolescents starting from diffusion move through more than two identity statuses to reach achievement. We elaborate further on these issues in the discussion.

Age differences. In the second step of LTA modeling, we did not find differences between age groups in rate of change into and out of identity statuses. Table 1, second and third upper panels, summarizes this similar and regular identity change for both cohorts. The table also shows systematic cohort differences in the prevalence of the statuses in Waves 1–5. In all waves, the number of diffusions, moratoriums, and SMs is higher in the younger age group, whereas the number of early closures/closures and achievers is lower. There are a few exceptions to this general pattern: In Waves 3 and 4, the prevalence of moratoriums was lower, and in Waves 4 and 5, the prevalence of early closures/closures was higher in the younger age group. We applied Bayesian model selection to evaluate which of three alternative models of the prevalence of the identity statuses in Waves 1 and 5 in both cohorts provided the best fit to the data. Model 1 assumed no difference in prevalence between the cohorts, whereas Model 2 assumed a higher prevalence of D, M and SM in early-to-middle adolescence and a higher prevalence of ECC and A in middle-to-late adolescence. Model 3, the unconstrained model, did not specify the distribution of statuses across cohorts. Table 2 presents the findings. The BFs imply that Model 1 is 250 times less likely than Model 3, and that Model 2 is 211.50 times more likely than Model 3. Moreover, Model 2 is 52,875 times as likely as Model 1. PMPs of Models 1, 2, and 3 are < .001, .99, and < .01, respectively. Replication of the same Bayesian models for Waves 2, 3, and 4 revealed similar differences between age groups. We do not include a full report of these models to save space. These findings show that the middle-to-late adolescents are generally in more "progressive" identity statuses than the early-to-middle adolescents.

Finally, Table 1 makes clear that the patterns of increases and decreases in identity status memberships unfold quite systematically from early-to-middle and middle-to-late adolescence. Diffusion decreases from 9.3% to 6.6% in early-to-middle adolescence, and from 7.9% to 5.5% in middle-to-late adolescence. For moratorium, the percentages decrease from 17.2 to 14.2 and from 14.4 to 13.3, and for SM the percentages decrease from 7.6 to 2.5 and from 2.1 to 0, whereas the percentages for achievement rise from 16.6 to 21.5 and from 21.8 to 26.2. Percentages for the ECC status rise from 49.3 to 55.3 in the younger cohort but are fairly stable in the older cohort (53.8%–55.1%). So, for diffusion, moratorium, and SM, a systematic decrease is evident from early to late adolescence; for achievement, we observe a systematic increase; and for the ECC identity status, there appears to be an increase followed by stabilization at a relatively high level.

Gender differences. In the second step of LTA modeling, we found gender differences in identity status transitions. Inspection of the separate 1-year transition tables indicated the general pattern of identity transitions to be the same for males and females. The seven primary results that we found for the transitions in the whole sample appeared to generalize across gender. However, one systematic gender difference appeared: In four of the five statuses (M, SM, ECC, and A), 1-year stability was more likely in females than in males—percentage differences were 1, 12, 2, and 4, respectively. To interpret this finding, we estimated two sets of Bayesian models on the two lower panels of Table 1. In the first set, we evaluated gender differences in the prevalence of identity statuses in Waves 1 and 5. Model 1 assumed no difference in prevalence between males and females, whereas Model 2 assumed a higher prevalence of D, M, and SM in males and a higher prevalence of ECC and A in females. Model 3, the unconstrained model, did not specify any distribution of the statuses for either gender. Table 2 presents the findings. The BFs indicated that model 1 is approximately 100 times less likely than Model 3, and that Model 2 is about 510 times more likely than Model 3. Moreover, Model 2 was 51,073 times more likely than Model 1. PMPs of Models 1, 2, and 3 were < .001, .99, and < .001, respectively—and as a result, we retained Model 2. Replication of the same Bayesian models for Waves 2, 3, and 4 revealed the same patterns of differences between males and females. We include no full report of the models to save space. These findings show that females are generally classified into more "progressive" identity statuses than males.

In the second set of Bayesian models, we evaluated three alternative models of gender differences in increases and decreases of identity status membership between Waves 1 and 5. Model 1 assumed no gender differences in likelihood of increase or decrease in the identity statuses over time, whereas Model 2 assumed a larger likelihood of decrease of D, M, and SM in males than in females and a larger likelihood of increase of ECC and A in males than in females. Model 3, the unconstrained model, did not specify any pattern of increase or decrease of identity statuses for either gender. Table 2 presents the findings from these model comparisons. The BFs imply that Model 1 was 250 times less likely to represent the data adequately than Model 3, and that Model 2 was about 5 times more likely than Model 3. Moreover, Model 2 was 1,322.5 times more likely to represent the data than Model 1. PMPs for Models 1, 2, and 3 were < .001, .84, and .16, respectively. These findings show that decreases in D, M, and SM, and increases in ECC and A, are more likely in males than females. Taken together, these findings indicate that females are in more advanced and more stable identity statuses than males, but that males may catch up during later adolescence.

Conclusion. Figure 2 summarizes the main findings of the final LTA model. The figure incorporates only those identity transitions with 4-year probabilities of .10 or above. The figure is based on an additional Bayesian model selection, in which we constrained the probabilities of the 4-year transitions below .10 (see Table 3) to be 0 and the probabilities of the other transition to be greater than 0. This model was clearly superior to a model in which the transitions were allowed to vary freely; PMPs of both models were .99 and .01, respectively. A comparable model comparison of the probabilities of the 1-year transitions replicated these findings. The figure clearly shows that progressive identity transitions outnumber regressive transitions. Identity progression is represented by seven transitions: D → M, D → EC, M → C, M → A, SM → C, SM → A, and EC → A. Identity regression is represented by three transitions: A → M, A → C, and SM → M.

Discussion

Identity Formation: Progression and Stability

This study was designed to test the extent to which identity status is a stable individual disposition or whether it changes over time. Our findings revealed a steady increase of A and ECC, and a steady decrease of D, M, and SM, in both cohorts. These findings support the first assumption of Waterman's (1982) developmental hypothesis and

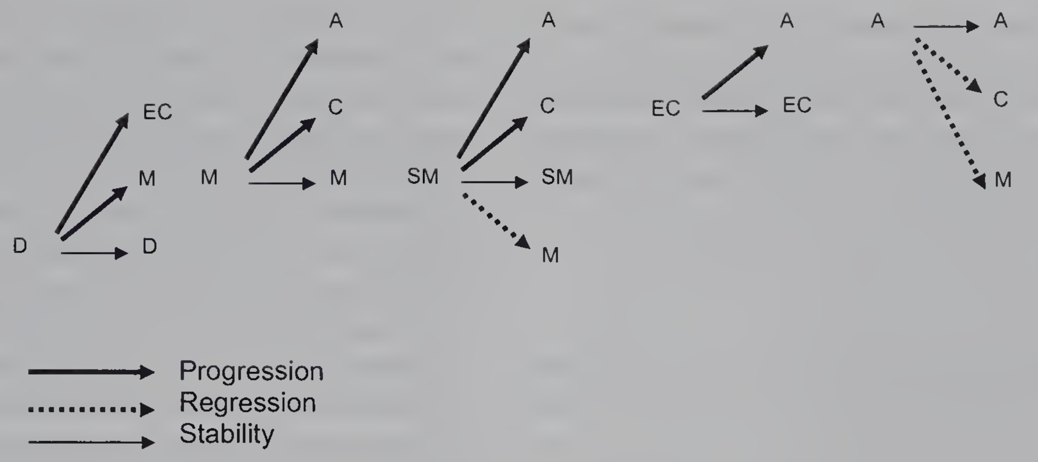


Figure 2. Transitions of identity progression, regression, and stability.
Note. D → EC, D → M, and M → A were hypothesized by Waterman (1982) as progressive transitions, and A → M as regressive transition. Percentages of the number of participants that changed statuses between Waves 1 and 5 can be found in the right-hand side of Table 3 (e.g., .11 signifies that 11% of the participants who were in diffusion in Wave 1 moved to moratorium in Wave 5).

document identity progression: growth of the high-commitment statuses (ECC and A), decrease in the number of adolescents who do not address identity issues (D), and decrease in the proportion of adolescents who are in the process of finding their identity (M and SM). Taken together, these findings clearly demonstrate identity maturation during adolescence, and they converge with the results of Klimstra et al. (2009), who have provided evidence for systematic personality maturation between early and late adolescence. Support for Waterman's second hypothesis was also substantial, but less consistent: we found support for three of the four progressive identity transitions that Waterman hypothesized, notably, D to EC, D to M, and M to A, and for one of the four regressive transitions he proposed: notably A to M. On the other hand, we also found substantial support for the individual difference perspective: Sixty-three percent of the adolescents remained in the same identity status in Waves 1 and 5. This percentage is remarkably similar to the 59% found in the longitudinal identity status studies reviewed in the introduction. So our study replicates earlier findings on change and stability of identity. We offer two interpretations for the high stability of identity status that we found. The first interpretation assumes that changes in identity may be more prevalent in emerging adulthood than in adolescence due to more frequent and intense consideration of adult roles. Some support for this position has been obtained in a recent meta-analysis of personality change by Roberts, Walton, and Viechtbauer (2006), who concluded that personality change is much more prevalent in young adulthood than in adolescence. The second assumption states that the universality of the identity conflict and subsequent identity formation has been wrongly and too commonly assumed. Support for this interpretation has been found in the review by Kroger (2007), who showed across cross-sectional and two-wave longitudinal studies no more than 25% of participants to be in moratorium. Future longitudinal studies from early adolescence into emerging adulthood are needed to more thoroughly test these alternative interpretations.

Another major finding is that Marcia's (1966) original four statuses—achievement, moratorium, foreclosure (here ECC), and diffusion—indeed emerged empirically as identity statuses at all five waves, along with a new status: searching moratorium. Our findings strongly suggest that SM is an early and middle adolescent status that disappears in late adolescence. So, in late adolescence (corresponding to the age group that Marcia used in his own research), we empirically extracted Marcia's four identity statuses. It is also of interest to note that the five-wave LTA model was characterized by high classification accuracy ($E = .85$).

Interestingly, we found the ECC status to be the most prevalent in both cohorts. A majority (80%) of the 725 adolescents who were in this status in Wave 5 can be considered to be early closures: They stayed in this status in all waves ($n = 530$) or made the transition from diffusion to early closure ($n = 53$). They are labeled as early closures because they have not considered identity alternatives and have continued to maintain relatively strong commitments over time. A minority (20%) of the adolescents with an ECC profile transition from moratorium, SM, or achievement. They have considered identity alternatives and therefore cannot be defined as early closures. We have labeled them as closures. At present, they possess relatively strong commitments and maintain them in an automatic way. They did not, however, always do so. We will discuss this issue further in the section on limitations and suggestions for further research.

Identity Transitions

Figure 2 summarizes the main findings of the final LTA model. The transitions depicted in this table are predominantly progressive. Identity progression is represented by seven transitions: D → M, D → EC, M → C, M → A, SM → C, SM → A, and EC → A. Identity regression is represented by three transitions: A → M, A → C, and SM → M. The seven progressive transitions indicate different changes in the identity configurations of adolescents: starting to think about alternative commitments (D → M), making stronger commitments (D → EC), making stronger or much stronger commitments along with decreases in considering alternative commitments (M → C and M → A, respectively), making much more secure (SM → C) or much more secure and active (SM → A) commitments, and making more active and strong commitments (EC → A). The three regressive identity transitions indicate discarding commitments and starting to consider alternative ones (A → M), moving from strong and active commitments to more rigid and less active commitments (A → C), and discarding commitments while maintaining high levels of reconsideration (SM → M).

The analyses of status transitions imply that the high-commitment statuses, A and ECC, are more likely than the other statuses to serve as endpoints of identity formation in adolescence. Stability of A and ECC is very substantial, and recent studies have documented that both statuses show more positive profiles of psychosocial adjustment compared to diffusions or moratoriums. Achievers and adolescents in foreclosure or ECC tend to be characterized by relatively low levels of depression (Meeus, 1996), anxiety (Berman, Weems, & Stickle, 2006), substance use (Luyckx, Goossens, Soenens, Beyers, & Vansteenkiste, 2005), aggression (Crocetti, Rubini, Luyckx, et al., 2008), and relatively high levels of emotional stability and self-esteem (Luyckx et al., 2005). In contrast to achievement and ECC, SM and diffusion appear to be almost exclusively transitional statuses, given that many adolescents move out of these statuses and few (if any) move into them.

The transitional analyses suggested only a limited amount of change. Few participants shifted from diffusion into achievement or from achievement to diffusion (Table 3). Secondly, analyses of the identity trajectories revealed that the majority of adolescents who change identity status across 5 years make only one transition. This makes clear that changes in identity status tend to be decisive, and that there is an extremely low probability of additional identity status transitions.

The analyses of identity transitions and trajectories shed new light on moratorium and reinforce the distinction between moratorium and SM. First, our findings particularly underscore the transitional nature of SM, as adolescents move primarily out of this status. In fact, there were no SMs in the last waves for the middle-to-late adolescent cohort. On the other hand, our analyses showed moderate stability among adolescents in moratorium. This finding suggests that a considerable number of moratoriums might be unable to move out of this unstable identity configuration, and should be considered as "characterological moratoriums," as Côté and Schwartz (2002, p. 584) have suggested. An important implication of Côté and Schwartz's suggestion is that, for characterological moratoriums, back-and-forth movement between moratorium and achievement, as has been found to occur in adulthood for some people (Stephen, Fraser, & Marcia, 1992), is improbable as an identity trajectory in adolescence. However, moratorium and SM may nonetheless function as transitory statuses. When adolescents make more than one identity transition, they tend to pass through moratorium or SM. "Classical" moratorium, however, may also be characterological, whereas SM appears to be exclusively transitory. In addition, our findings show that SM offers a better starting point to reach achievement than does moratorium: 32% versus 22% (4-year transition probabilities), respectively.

As noted earlier, our findings partially support the second assumption of Waterman's (1982) developmental hypothesis. Waterman hypothesized that progressive identity development would occur through the transitions D → EC, D → M, EC → M, and M → A. We found evidence for the progressive transitions D → EC, D → M, and M → A. The explanation for the nonoccurrence of the transition EC → M is that, as noted before, the early closure status functions very often as the final identity status observed in adolescence. Adolescents in this status do not think a lot about their present commitments and are not active in searching for alternative commitments. Therefore, they do not appear likely to give up their commitments and consider adopting alternative life choices. With regard to regressive transitions, we found support for one of the four regressive identity status transitions that Waterman hypothesized: from achievement to moratorium. Achievers are adolescents with the highest level of in-depth exploration. They are very active in processing information about their commitments. This orientation may give rise to loss of present commitments and search for alternative commitments if one's present commitments are deemed unsatisfactory (Luyckx et al., 2006). We did not find support for the three other regressive transitions that Waterman hypothesized: ECC → D, M → D, and A → D. Our findings clearly suggest that it is almost impossible to go back to a state of disinterest in identity work (D) once individuals have held strong commitments (ECC and A) or have been active in considering alternative commitments (M). In sum, consistent with earlier research (Berzonsky & Adams, 1999; Kroger, 2007; Van Hoof, 1999), we found more progression than regression. At the same time, our findings also make clear that regression in identity is something that cannot be ignored.

Gender Differences

We found considerable gender differences in patterns of identity formation. As expected, females were more likely to be achieved, and less likely to be diffused, than males. In addition, we also found that females were more likely to be classified into the ECC status, and less likely to appear in both moratoria, than males. These findings suggest that, at least at present, females may be "further ahead" of males when overall identity is measured as a combination of interpersonal and ideological (especially educational) domains in adolescence and the early part of emerging adulthood. The explanation might be that Dutch females combine their classic stronger interpersonal commitments (Meeus & Deković, 1995) with stronger educational commitments, given that girls at present often outperform boys in school (Statistics Netherlands, 2008b, 2008c). Our findings are consistent with results of recent studies in the United States that have reported gender differences in interpersonal identity (Lewis, 2003), ideological identity (Schwartz & Montgomery, 2002), and overall identity (Guerra & Braungart-Rieker, 1999). We also found, however, that males tend to "catch up" during adolescence. This is consistent with the review by Kroger (1997), and suggests that earlier physical and cognitive maturation in girls may account for some of this pattern. Recent studies have shown that girls reach puberty between 1 and 2 years earlier than boys (Beunen et al., 2000) and that, in early adolescence, girls tend to be up to a full year ahead of boys in several aspects of brain development (Colom & Lynn, 2004; Giedd et al., 1999). Therefore, girls might reach the mature identity statuses earlier than boys, whereas boys catch up during adolescence.

Limitations and Suggestions for Further Research

A number of limitations of this study warrant discussion. The first limitation involves our sole reliance on self-report questionnaires. Although questionnaires are the most appropriate instruments by which to gather information on subjective processes, such as identity development, the biases involved in self-reports may have come into play. Future research could try to overcome these biases by focusing on the micro processes underlying identity development. This could be done by tapping identity on a day-to-day basis.

A second limitation of our study is its descriptive nature. We did not test explanations of identity progression during adolescence—for example, why certain identity transitions have a higher probability than others. Given that identity status transitions in adolescence seem to be quite decisive, future research should try to specify the conditions that predict the timing of these transitions. Longitudinal designs that include a focus on the link between identity transitions and life transitions, that is transitions in the educational and occupational career and in the formation of intimate relationships, may be a fruitful option here.

Although our model is conceptually distinct from Marcia's (1966) model in two respects, the findings of our study are remarkably similar to those of studies using Marcia's paradigm. Notably, we found similar percentages of change and stability of identity statuses, moratorium to be among the least stable statuses over time, and that the high-commitment statuses are associated with the most positive adjustment profile. Notably, different is the very high prevalence of ECC in our study as compared to that of foreclosure in the earlier studies using Marcia's paradigm. Obviously, this finding requires replication and expansion, for instance by including more domains in an overall identity measure, such as dating relationships, work, religion, and politics. It is also important for future research to examine whether the high prevalence of ECC in our study is due to our use of reconsideration instead of Marcia's original measure of exploration in breadth. Adding a measure of exploration in breadth to the U-MICS could clarify this.

Despite these limitations, this study has contributed significantly to our understanding of the process of identity formation over time. It is the first five-wave study of a broad-range sample of early-to-middle and middle-to-late adolescents to show how identity develops between the ages 12–20 and to elucidate which identity transitions are most likely to characterize these changes. The makeup of our sample suggests that our findings may be generalizable to individuals who are in various types of education during adolescence. Findings of our study may be less generalizable to adolescents who enter the labor force very early and to adolescents from ethnic minority groups. We also clearly demonstrated that statuses with a very clear resemblance to those from Marcia's (1966) model emerged in all waves in both early-to-middle and middle-to-late adolescence. It is hoped that these findings inspire more longitudinal research on identity development.

References

Adams, G. R., & Fitch, S. A. (1982). Ego stage and identity status development: A cross-sequential analysis. *Journal of Personality and Social Psychology*, *43*, 574–583.

Adams, G., & Jones, R. (1983). Female adolescents' identity development: Age comparisons and perceived child-rearing experience. *Developmental Psychology*, *19*, 249–256.

Archer, S. (1982). The lower age boundaries of identity development. *Child Development*, *53*, 1551–1556.

Berman, S. L., Weems, C. F., & Stickle, T. R. (2006). Existential anxiety in adolescents: Prevalence, structure, association with psychological symptoms and identity development. *Journal of Youth and Adolescence*, *35*, 303–310.

Berzonsky, M., & Adams, G. (1999). Reevaluating the identity status paradigm: Still useful after 35 years. *Developmental Review*, *19*, 557–590.

Beunen, G., Thomis, M., Maes, H. H., Loos, R., Malina, R. M., Claessens, A. L., et al. (2000). Genetic variance of adolescent growth in stature. *Annals of Human Biology*, *27*, 173–186.

Bollen, K. (1989). *Structural equation modeling with latent variables*. New York: Wiley.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, *36*, 75–82.

Côté, J., & Schwartz, S. (2002). Comparing psychological and sociological approaches to identity: Identity status, identity capital, and the individualization process. *Journal of Adolescence*, *25*, 571–586.

Crocetti, E., Rubini, M., Luyckx, K., & Meeus, W. (2008). Identity formation in early and middle adolescents from various ethnic groups: From three dimensions to five statuses. *Journal of Youth and Adolescence*, *37*, 983–996.

Crocetti, E., Rubini, M., & Meeus, W. (2008). Capturing the dynamics of identity formation in various ethnic groups: Development and validation of a three-dimensional model. *Journal of Adolescence*, *31*, 207–222.

Dellas, M., & Jernigan, L. P. (1987). Occupational identity status development, gender comparisons, and internal-external control in first-year air force cadets. *Journal of Youth and Adolescence*, *16*, 587–600.

Erikson, E. (1968). *Identity, youth and crisis*. New York: Norton.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, 2, 861–863.

Grotevant, H. (1987). Toward a process model of identity formation. *Journal of Adolescent Research*, *2*, 203–222.

Guerra, A., & Braungart-Rieker, J. (1999). Predicting career indecision in college students: The role of of identity formation and parental relationship factors. *Career Development Quarterly*, *47*, 255–266.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, *44*, 457–467.

Klimstra, T. A., Hale, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. (2009). Maturation of personality in adolescence. *Journal of Personality and Social Psychology*, *96*, 898–912.

Kroger, J. (1988). A longitudinal study of ego identity status interview domains. *Journal of Adolescence*, *11*, 49–64.

Kroger, J. (1997). Gender and identity: The intersection of structure, content, and context. *Sex Roles*, *36*, 747–770.

Kroger, J. (2007, March). *The status of identity: Developmental perspectives*. Paper presented at the annual meeting of the Society for Research on Identity Formation, Washington, DC.

Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research*, *16*, 123–138.

Lewis, H. (2003). Differences in ego identity among college students across age, ethnicity, and gender. *Identity*, *3*, 159–189.

Luyckx, K., Goossens, L., & Soenens, B. (2006). A developmental contextual perspective on identity construction in emerging adulthood: Change dynamics in commitment formation and commitment evaluation. *Developmental Psychology*, *42*, 366–380.

Luyckx, K., Goossens, L., Soenens, B., Beyers, W., & Vansteenkiste, M. (2005). Identity statuses based upon four rather than two identity dimensions: Extending and refining Marcia's paradigm. *Journal of Youth and Adolescence*, *34*, 605–618.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, *3*, 551–558.

Markstrom, C., & Marshall, S. (2007). The psychosocial inventory of ego-strengths: Examination of theory and psychometric properties. *Journal of Adolescence*, *30*, 63–79.

Markstrom, C., Sabino, V., Turner, B., & Berman, R. (1997). The psychosocial inventory of ego-strengths: Development and validation of a new Eriksonian measure. *Journal of Youth and Adolescence*, *26*, 705–732.

Meeus, W. (1996). Studies on identity development in adolescence: An overview of research and some new data. *Journal of Youth and Adolescence*, *25*, 569–598.

Meeus, W., & Deković, M. (1995). Identity development, parental and peer support in adolescence: Results of a national Dutch survey. *Adolescence*, *30*, 931–945.

Meeus, W., Iedema, J., Helsen, M., & Vollebergh, W. (1999). Patterns of adolescent identity development: Review of literature and longitudinal analysis. *Developmental Review*, *19*, 419–461.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles. Author.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric group based approach. *Psychological Methods*, *4*, 139–157.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.

Nylund, K., Muthén, B., Nishina, A., Bellmore, A., & Graham, S. (2006). *Stability and instability of peer victimization during middle school: Using latent transition analysis with covariates, distal outcomes, and modeling extensions*. Unpublished manuscript.

Reinecke, J. (2006). Longitudinal analysis of adolescent's deviant and delinquent behavior. *Methodology*, *2*, 100–112.

Reinke, W., Herman, K., Petras, H., & Ialongo, N. (2008). Empirically derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology*, *36*, 759–770.

Roberts, B., Walton, K., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life-course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*, 1–25.

Schwartz, S., & Montgomery, M. (2002). Similarities or differences in identity development? The impact of acculturation and gender on identity process and outcome. *Journal of Youth and Adolescence*, *31*, 359–372.

Statistics Netherlands. (2008a). *Statline: Bevolking naar Herkomst en Generatie*. Voorburg, Netherlands: Statististics Netherlands.

Statistics Netherlands. (2008b). *Statline: Mbo; leerlingen en geslaagden*. Voorburg, The Netherlands: Statististics Netherlands.

Statistics Netherlands. (2008c). *Statline: Vo; leerlingen en geslaagden*. Voorburg, Netherlands: Statistics Netherlands.

Stephen, J., Fraser, E., & Marcia, J. E. (1992). Moratorium-achievement (Mama) cycles in lifespan identity development: Value orientations and reasoning system correlates. *Journal of Adolescence*, *15*, 283–300.

Van Hoof, A. (1999). The identity status field re-reviewed: An update of unresolved and neglected issues with a view on some alternative approaches. *Developmental Review*, *19*, 497–556.

Waterman, A. S. (1982). Identity development from adolescence to adulthood: An extension of theory and a review of Research. *Developmental Psychology*, *18*, 341–358.

Waterman, A., Geary, P., & Waterman, C. (1974). Longitudinal study of changes in ego identity status from the freshman to the senior year at college. *Developmental Psychology*, *10*, 387–392.

Waterman, A., & Goldman, J. (1976). A longitudinal study of ego identity development at a liberal arts college. *Journal of Youth and Adolescence*, *5*, 361–369.

Waterman, A. S., & Waterman, C. K. (1971). A longitudinal study of changes in ego identity status during the freshman year at college. *Developmental Psychology*, *5*, 167–173.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1582–1597

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning

Deborah L. Linebarger and Jessica Taylor Piotrowski
University of Pennsylvania

Educational TV has been consistently linked to children's learning. In this research, educational TV characteristics were identified, coded, and tested for their influence on children's program-specific comprehension and vocabulary outcomes. Study 1 details a content analysis of TV features including a program's macrostructure (i.e., narrative or expository) and learning strategies embedded in the macrostructure that support learning in print-based contexts. In Study 2, regression analyses were used to predict outcomes involving 71 second and third graders (average age = 7.63 years). Strategies were categorized as organizing, rehearsing, elaborating, or affective in function. Outcomes were uniformly higher for narrative macrostructures. Strategies used in narratives predicted relatively homogenous relations across outcomes, whereas strategies in expositories predicted quite heterogeneous relations across outcomes.

Since the introduction of *Sesame Street* in 1969, producers, educators, and researchers have worked together and in parallel to create and evaluate the impacts of educational TV. When programs are created using entertaining formats guided by developmental theory, researchers document increases on various outcomes including school readiness (e.g., Wright et al., 2001), problem solving (e.g., Crawley, Anderson, Wilder, Williams, & Santomero, 1999), and literacy (e.g., Linebarger, Kosanic, Greenwood, & Doku, 2004). There is growing evidence that embedding content with certain learning strategies supports children's processing of that content. Showing a word that breaks into individual sounds and then reassembles into the word helps early readers acquire phonemic awareness (e.g., *mm*, *aa*, *pp* when blended together forms *map*; Linebarger et al., 2004; Uchikoshi, 2005) while pairing visual referents with verbal labels or using intonational cues to emphasize content helped children learn more (Fisch, 2004).

Learning involves a relatively permanent change in behavior or knowledge. The most straightforward theories of learning from TV are content based and involve the acquisition, encoding, storage, and retrieval of information. These include behavior change (e.g., social cognitive theory; Bandura, 2001) and information-processing theories (capacity model; Fisch, 2004) that, in their broadest forms, predict that knowledge results from cognitively processing televised content. When children can process and comprehend visual and verbal TV symbols, they will be able to organize, store, and relate this information to prior knowledge. Acquired knowledge and behavior are then used to influence a child's actions.

Focusing attention on particular televised stimuli is dependent on an interaction among interest, ability, and characteristics inherent in stimuli. Becoming sufficiently aroused and oriented to a stimulus while blocking out other, irrelevant stimuli (Herrman, Yoder, Gruneberg, & Payne, 2006) is necessary for learning. Children direct maximal attention to stimuli perceived to be moderately novel and of intermediate complexity (Rice, Huston, & Wright, 1982; Roller, 1990). Declining attention occurs when the material is overly familiar and easy to process, or, conversely, too novel and complex. One way to

---

This project was supported by two U.S. Department of Education (DOE) grants (H029D60040; H327A990082) and a third U.S. DOE cooperative agreement (U295A050003). Please note, these contents do not necessarily represent the policy of the U.S. DOE and you should not assume endorsement by the federal government. Portions of the research were presented at the annual meeting of the Office of Special Education Projects, U.S. DOE, Washington, DC, June 2002, and at the biennial meeting of the Society for Research in Child Development, Tampa, FL, April 2003. Special thanks to the staff and participants involved in this project including Dr. John C. Wright of the University of Texas at Austin who was a coinvestigator on this project until his death in 2001. We are indebted to his many contributions. His guidance and mentoring will be missed by all those who knew him or benefited from his scholarly brilliance.

Correspondence concerning this article should be addressed to Deborah L. Linebarger, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104. Electronic mail may be sent to dlinebarger@asc.upenn.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0019

boost attention and support content retention is to insert learning strategies that support information-processing efforts. There is considerable variability in the degree of instructional support provided by programs, ranging from those that rely heavily on viewers to generate and apply their own strategies to those that use strategies to explicitly structure viewers' processing in an effort to stimulate learning (Smith & Ragan, 2005).

Recently, educational media began creating content to support the development of early literacy skills. The purpose of this study was to evaluate a set of learning strategies hypothesized to support children's literacy abilities in both narrative and expository programs. To operationalize and evaluate these learning strategies, two studies were conducted. In Study 1, various literacy-promoting learning strategies were culled from print and other literacy-related research and then coded for their presence in children's educational TV. All programs used one of two program types or macrostructures: narrative or expository. In Study 2, learning strategies and macrostructures were used to predict program-specific comprehension and vocabulary.

Study 1

A number of learning strategies that support print-based literacy were identified from several sources (National Institute of Child Health and Human Development, 2000; Snow, Burns, & Griffin, 1998). Because research on the effects of TV on young children has established the crucial role that content plays, because programs with well-designed curricula produce stronger academic and social effects (e.g., Anderson, Huston, Schmitt, Linebarger, & Wright, 2001), and because strategies typically work across different media, it was appropriate to identify learning strategies used in other contexts that might also be found in educational TV. Via an iterative process of review and coding, three key sets of strategies were identified: affective strategies (i.e., literacy environment), comprehension strategies, and vocabulary strategies. Each is reviewed next.

Literacy Environment

Programs may explicitly or implicitly create supportive literacy environments by depicting characters who use print or make positive comments about literacy as well as by displaying print onscreen. Making a decision to invest attention and cognitive resources to televised content is dependent on a child's personal interest, the stimuli characteristics, and the interaction between the two. Personal interest develops when the child's perception of the rewards for engaging in the modeled set of behaviors is strong (Schiefele, 1998). According to Bandura's (2001) social cognitive theory, children observe and imitate persons found in their everyday worlds, including characters on TV, particularly when those characters are interesting and personally relevant. Stimuli characteristics also generate interest and attention when they help children believe they could engage in and successfully replicate onscreen behavior or that literacy is valued and respected. When questioned about their competence in learning from TV and print, the competence scores of poor at-risk second graders were rated more highly for learning from TV (Linebarger, 2001). Motivation levels increased when instruction was based on popular TV programs (Szabo & Lamielle-Landy, 1981). Greater motivation results in deeper processing of content (Renninger, 1992) that, for TV, was found to transfer to higher fourth-grade reading achievement (Szabo & Lamielle-Landy, 1981). Two longitudinal studies involving preschoolers also provide support: Watching educational TV as a preschooler predicted better school readiness skills at school entry (Wright et al., 2001) and more leisure book reading, greater academic self-concept, and higher English grade point averages when a teen (Anderson et al., 2001).

Onscreen print should contribute to literacy skills when that print is a natural part of the plot (e.g., characters write notes or grocery lists; Fisch, 2004; Linebarger et al., 2004). Viewers are given opportunities to practice reading onscreen print with visual (e.g., print grows and shrinks) and verbal supports (e.g., intonation cues). Receiving stimuli in multiple modalities strengthens the connections among these stimuli, resulting in efficient encoding and later retrieval (Sadoski & Paivio, 2001). Incorporating print as an extension of character or program events links it to the main theme, making it more meaningful and easier to remember (Fisch, 2004).

Comprehension Strategies

Constructing meaning from content requires prior knowledge about that content and explicit links between that knowledge and new content. Strategies that support comprehension include activation of prior knowledge (via advance organizers or connections to real life) and repetition and review of the content after it has been presented (Michel & Roebers, 2008). Comprehension is also supported when new information is limited (Reinking & Schreiner, 1985) or when readers are directed to make predictions about upcoming content (Taylor, Peterson, Pearson, & Rodriguez, 2002). Using comprehension strategies should help children remember more of an episode as well as learn the underlying educational content in that episode.

Vocabulary Strategies

Children are able to learn specific words from TV as well as transfer that knowledge to more generalized vocabulary gains (Rice & Woodsmall, 1988; Wright et al., 2001); however, these associations hold only for educational TV (e.g., Patterson, 2002). Naigles and Mayeux (2001) concluded that when designed appropriately using strategies that support word learning in print-based contexts, educational TV helps vocabulary acquisition. TV affords a unique opportunity to present words and concepts using a combination of images, verbal description, and sound effects. When supporting text is simple, more challenging content can be inserted. The simple text requires little processing and, as such, directs attention to the central content and away from incidental content. The rate of aural narration also affects what and how much is comprehended. Pairing static text with narration (e.g., books on tape) increased reading vocabulary and definition knowledge when narration rates were similar to or just above the child's reading rate (McMahon, 1983). Televised and audio-taped book narration rates average 141 words per minute (wpm), children's educational TV averages 124 wpm, and cartoons average 125 wpm (Jensema, McCann, & Ramsey, 1996). Conversation rates are between 150 and 175 wpm (Taylor, 1964). While first to third graders are able to comprehend content at these speeds, it begins to decline as speeds reach and surpass 175 wpm (Barron, 2004). Lower rates are required when children have reading difficulties or when content is especially complex (e.g., science content is more challenging than prosocially themed stories at faster narration rates; Barron, 2004). While these studies do not exactly mirror the process of viewing and comprehending TV, recent research finds that similar skills underlie processing of different media. Comprehension of televised stories at age 6 predicted reading comprehension at age 8 (Kendeou et al., 2005).

Program Type: Narrative Versus Expository Macrostructures

In addition to learning strategies, the broader macrostructure used to deliver content affects learning. While different narrative macrostructures exist (e.g., traditional story book, story within a story; Linebarger & Piotrowski, 2009), the foundational elements used in most narratives are similar (e.g., setting, character, goals, resolution). In contrast, expository macrostructures present information in varied ways depending on the purpose of the information (e.g., compare and contrast, listing and sequencing, cause and effect; Duke & Kays, 1998). Embedded within both macrostructures are learning strategies that support processing by reducing cognitive load and making content accessible and salient. It is unclear whether these strategies function similarly across macrostructures or whether their influence is dependent on, shaped by, or interacting with the macrostructure. It is likely that these strategies function in both ways.

The skills used to interpret content and those that develop from repeated exposure to content differ by the macrostructure used. Reading expository books contributes to conceptual knowledge and vocabulary development, stronger comparative and contrastive abilities, and the use of classificatory language structures (Duke & Kays, 1998). In contrast, exposure to narrative stories supports language and literacy skills, story knowledge, and later school success (e.g., Nord, Lennon, Liu, & Chandler, 1999; Scarborough & Dobrich, 1994).

Television programs also use these two macrostructures although little research has specifically focused on the effects of these macrostructures. Narratives contribute to language and literacy skill acquisition (Kendeou et al., 2005; Linebarger & Piotrowski, 2009; Linebarger & Walker, 2005). Watching narratives over an extended period helped preschoolers augment story knowledge and print-based narrative skills compared with children who viewed an expository or no program (Linebarger & Piotrowski, 2009). Story knowledge and narrative abilities develop from the ability to identify the meaning of a particular story. When a program uses a narrative macrostructure, children should be better able to recognize story parts (e.g., goals, characters, setting) and, more important, the deeper educational content (Fisch, 2004).

The utility of expositories, comparisons between these programs and narratives, and comparisons between these programs and comparable expository printed texts, have not been directly evaluated.

There is, however, sufficient evidence that children learn content from expositories and are able to transfer that content to more generalized knowledge gains (e.g., Fisch, 2004). Questions remain about what content is best learned through which macrostructure and whether particular learning strategies support or interfere with learning that content.

Method

Sample

Six 30-min programs previously or currently airing on the Public Broadcasting Service (PBS) channel were selected for inclusion in this study. Curriculum documents and promotional materials developed by each of the programs indicated that these programs were designed for children in the target age group (i.e., 7- to 9-year-olds). PBS programs were selected because (a) PBS is widely available and free, (b) not all children in the study had access to cable (e.g., for programs on Nickelodeon or Disney), and (c) there were no commercials on PBS. In addition to airing on PBS and targeting second- and third-grade children, an equal number of narrative and expository macrostructures were included. Narratives used a set of story events containing dialogue, characters, setting, goals, and outcomes to deliver prosocial or academic content. Expositories used multiple vignettes and program hosts to educate, inform, or describe a particular topic. As the goal of this research was to make more general statements about macrostructures and embedded learning strategies (vs. claims about particular programs), selecting six different programs equally representing both macrostructures provided stronger tests of our hypotheses. *Arthur & Friends* (ART), *Magic School Bus* (MSB), and *Wishbone* (WB) were selected as narrative exemplars. *Reading Rainbow* (RR), *Zoom* (ZM), and *Kratts' Creatures* (KC) were selected as expository exemplars.

Coding Framework

Literacy environment. To evaluate whether the literacy environment depicted in an episode was conducive to learning, the following were coded: character's use of print (i.e., reading books, writing, using the computer; Print Use), character's positive statements about print (i.e., frequency; Positive Statements), and onscreen print (i.e., frequency that print onscreen lasted at least one second; Onscreen Print).

Comprehension strategies. Four comprehension strategies were coded. An episode's number of main segments was counted (i.e., segments introducing new information related to an overall theme; Main Segments). Pre- and postmain segments were coded as Prior Knowledge or Summary, respectively. Prior Knowledge was the duration in seconds that characters or a narrator spent directing viewers to access their own background knowledge. Summary was the duration in seconds when the information in a Main Segment was repeated exactly or paraphrased after its initial presentation. Characters' or narrator's use of questions and predictions related to upcoming content were scored as Questions/Predictions (i.e., frequency).

Vocabulary strategies. A variety of components related to vocabulary were coded including number of times target words repeated (i.e., Target Words), number of words per sentence (i.e., Sentence Length), the percentage of Dolch words used (i.e., Dolch Words; Dolch words represent 220 high-frequency words that comprise 50%–70% of kindergarten through third-grade primers; May, 1998), and the pace or rate of the aural narration (i.e., Narration Rate).

Reliability

Krippendorff's alpha (Hayes & Krippendorff, 2007) was calculated to examine intercoder reliability for frequency counts and durations. Frequency counts were tallied for Print Use, Positive Statements, Onscreen Print, and Questions/Predictions, while durations were computed for Prior Knowledge and Summary. All codes demonstrated strong intercoder reliability: Print Use = .99, Positive Statements = .96, Onscreen Print = .99, Prior Knowledge = .97, Summary = .98, and Questions/Predictions = .99.

The number of Main Segments was identified together with each coder prior to coding the other variables. Vocabulary codes were calculated using Concordance software (Watt, 2007). After entering episode scripts, the software counted the total number and frequency of all words (including Target Words and Dolch Words) and the average words per sentence (i.e., Sentence Length). Narration Rate was calculated by dividing the total words spoken by the episode air-time. Table 1 provides descriptive data, and Table 2 provides correlations among all variables.

Analytical Approach

To facilitate comparisons across codes and programs and balance the contribution of each strategy

Table 1

Means and Analysis of Variance (ANOVA) Results by Program (P) and Macrostructure (MS)

	Program means								
	Narrative			Expository					
Category codes	ART	MSB	WB	RR	ZM	KC	ANOVA, *F* (partial η^2)		
Literacy environment (LE) by program (P)							P	LE	P × LE
PU	2.0	7.5	22.0	1.5	6.0	14.0	3.67† (0.753)	0 (0)	1.16 (0.492)
PS	4.5	0.5	5.0	1.5	1.5	1.5			
OP	18.0	25.5	21.0	17.5	54.5	12.5			
Literacy environment (LE) by macrostructure (MS)							MS	LE	MS × LE
PU	12.67			14.00			0.95 (0.087)	0 (0)	0.09 (0.019)
PS	2.17			2.67					
OP	20.33			29.33					
Comprehension strategies (CS) by program (P)							P	CS	P × CS
MS	5.0	8.0	5.5	6.0	9.0	9.5	14.52** (0.924)	0 (0)	1.68 (0.584)
PK	14.0	399	194.0	409.5	129.5	353.0			
SM	111.5	417.5	156.5	1143.0	98.0	1131.0			
QP	66.5	85.5	22.0	63.5	46.5	33.0			
Comprehension strategies (CS) by macrostructure (MS)							MS	CS	MS × CS
MS	6.33			8.00			1.45 (0.126)	0 (0)	9.03** (0.772)
PK	316.17			225.50					
SM	557.33			461.83					
QP	71.83			33.83					
Vocabulary strategies (VS) by program (P)							P	VS	P × VS
TW	13.0	33.0	25.0	23.5	28.5	27.0	3.64† (0.752)	0 (0)	6.47*** (0.844)
NR	10.7	9.6	84.6	99.1	113.4	109.1			
SL	5.5	4.9	9.2	6.1	7.3	8.9			
%D	47.2	43.3	5.5	47.5	54.8	46.7			
Vocabulary strategies (VS) by macrostructure (MS)							MS	VS	MS × VS
TW	23.17			26.83			15.16** (0.602)	0 (0)	1.16 (0.303)
NR	96.82			102.39					
SL	5.50			8.45					
%D	46.04			50.48					

Note. Code/range: PU = Print Use (2–32 times); PS = Positive Statements (0–6 statements); OP = Onscreen Print (5–71 instances); MS = Main Segments (5–12 segments); PK = Prior Knowledge (100–627 s); SM = Summary (79–1273 s); QP = Questions and Predictions (9–96 times); TW = Target Words (11–36 repetitions); NR = Narration Rate (81.9–119.0 words spoken per minute); SL = Sentence Length (4.5–9.8 words per sentence); %D = Percent Dolch Words (42.4%–54.0%).
$^{\dagger}p < .10$. $^{**}p < .01$. $^{***}p < .001$.

to the overall models, all coded variables were *z*-score transformed. Six (programs) by *N* (category of coded variables) repeated measures analyses of variance (ANOVAs) with repeated measures on the categories (e.g., Literacy Environment, $N = 3$: Print Use, Positive Statements, Onscreen Print) were computed to examine code-related program differences. To evaluate whether macrostructures differentiated strategy use, ANOVAs comparing each type were computed. In these analyses, *z* scores were used to generate the *F* tests reported in Table 1. To ease interpretation, the original means are presented in Table 1. Because the sample size was low (i.e., $N = 6$) and to estimate practical significance, effect sizes are also reported in Table 1. All reported results are significant at alpha level $p < .05$ or better.

Results

Variability in Strategies Across Programs

Literacy environment. To test for differences in the depiction of the literacy environment across programs, a 6 × 3 (Programs × Literacy Environment) repeated measures ANOVA with repeated measures on the last factor was conducted. The literacy environment was marginally variable across programs ($p < .10$). RR contained the densest literacy environment followed by ART, ZM, KC, WB, and MSB.

Table 2

Correlations Among Literacy Environment, Comprehension Strategies, and Vocabulary Strategies

Variables	1	2	3	4	5	6	7	8	9	10	11
1. Print use	—										
2. Positive statements	.40***	—									
3. Onscreen print	−.12***	−.40***	—								
4. Main segments	−.39***	−.58***	.33***	—							
5. Prior knowledge	−.24***	−.23***	−.43***	−.16***	—						
6. Summary	−.09*	−.43***	−.41***	.16***	.55***	—					
7. Questions/predictions	−.21***	−.42***	.09*	.01	.12***	−.06	—				
8. Target words	−.24***	−.68***	.47***	.64***	.00	.22***	.03	—			
9. Narration rate	−.40***	−.31***	.36***	.67***	−.28***	.17***	−.04	.13***	—		
10. Sentence length	.12**	.31***	−.10**	.26***	−.18***	.09**	−.92***	.11**	.17***	—	
11. Dolch words	−.02	.25***	.57***	.19***	−.41***	−.42***	−.45***	−.06	.40***	.47***	—

*$p < .05$. **$p < .01$. ***$p < .001$.

Comprehension strategies. To evaluate whether comprehension strategies differed across programs, a 6 × 4 (Programs × Comprehension Strategies) repeated measures ANOVA with repeated measures on the last factor was computed. Comprehension strategies significantly varied across programs while the distribution and quantity of each was similar within programs. MSB contained the most comprehension strategies followed by WB, KC, ZM, ART, and RR.

Vocabulary strategies. To examine variation in the use of vocabulary strategies, a 6 × 4 (Programs × Vocabulary Strategies) repeated measures ANOVA with repeated measures on the last factor was computed. Vocabulary strategies varied marginally across programs. ZM contained the most, followed by KC, RR, WB, MSB, and ART. The distribution of strategies within programs also differed. The density of all four strategies was higher than average in ZM and lower than average in WB. Both KC and RR contained more strategies than average across three of the four codes (i.e., KC contained a lower than average percentage of Dolch Words while RR contained a slower than average Narration Rate). The reverse was true for ART and MSB; both contained fewer than average strategies for three of the four codes. ART's Narration Rate was faster than average while MSB repeated Target Words more often than average.

Macrostructure Differences

Literacy environment. To test if the literacy environment varied by macrostructure, a 2 × 3 (Macrostructure × Literacy Environment) repeated measures ANOVA with repeated measures on the last factor was computed. The literacy environment did not vary by macrostructure.

Comprehension strategies. The macrostructure was hypothesized to differentiate comprehension strategies such that narratives would use more effective comprehension strategies. This hypothesis was tested by computing a 2 × 4 (Macrostructure × Comprehension Strategies) repeated measures ANOVA with repeated measures on the last factor. This hypothesis was supported. The significant two-way (Macrostructure × Comprehension Strategies) interaction indicated that narratives used more supportive comprehension strategies and fewer comprehension obstacles compared with expositories (i.e., longer Prior Knowledge and Summary segments, more Questions/Predictions, fewer Main Segments than expositories).

Vocabulary strategies. Macrostructure was hypothesized to differentiate the use of vocabulary strategies such that expositories would use more effective vocabulary strategies. To test this hypothesis, a 2 × 4 (Macrostructure × Vocabulary Strategies) repeated measures ANOVA with repeated measures on the last factor was computed. This hypothesis received partial support. The significant two-way (Macrostructure × Vocabulary Strategies) interaction indicated that expositories contained more of all strategies, supportive (i.e., more repetitions of Target Words, greater percentage of Dolch Words) and unsupportive (i.e., longer Sentence Length, faster Narration Rates) when compared with narratives.

Discussion

Study 1 confirmed modest to high levels of variability in learning strategies across and within programs. Macrostructures provided a framework to which microlevel learning strategies were affixed. Significant variability by macrostructure for comprehension and vocabulary strategies was found while the literacy environment was depicted similarly for both.

Narratives contained fewer main segments, easier program text, and more time spent activating prior knowledge and summarizing program content. These strategies should reduce the cognitive resources needed to make sense of content by explicitly instructing viewers to access relevant background knowledge and by offering time to reflect on new content during summaries. Children, by virtue of their considerable narrative experience (e.g., 92% of the top 25 children's programs in 2003 used narratives, Nielsen Media Research, 2003; 90% of books in kindergarten through third grade use narrative formats, Stein & Trabasso, 1982), should be able to devote fewer resources to processing narrative macrostructures, freeing up cognitive capacity to generate inferences.

Conversely, expositories are organized to transmit information and, as a consequence, contained more unique content ideas and greater structural complexity than narratives. As a result, expositories tend to demand greater organizational and processing skills than is typically required for narratives (Best, Floyd, & McNamara, 2008). Both supportive and unsupportive vocabulary learning strategies were highest in expositories while comprehension strategies were used less frequently than narratives. It is unclear whether the supportive vocabulary strategies (i.e., target words, Dolch words) are more powerful than the unsupportive strategies (i.e., longer sentences, faster narration rates), leading to higher literacy scores or whether vocabulary complexity nullifies the supportive vocabulary strategies, leading to lower literacy scores. In contrast, the lack of explicit comprehension supports makes it likely that viewers will encounter difficulty linking their own knowledge (which, for at-risk children, is often limited) to the new information, resulting in diminished program comprehension.

Study 2

Research and theory assert that children's vocabulary knowledge and comprehension abilities will be higher for programs that depict a more positive literacy environment, that use more comprehension strategies, and that use more vocabulary strategies. Understanding program content results from a child's ability to recognize the macrostructure used to present that content as well as the microlevel strategies that define the organization of and elements within a particular media text. The microlevel strategies refer to how ideas are arranged and connected to one another and how visual or verbal cues signal key content. This is successfully accomplished when children are familiar with the purposes behind a particular macrostructure and, more specifically, with the strategies that act as cues for key information. Based on Study 1, it is hypothesized that each macrostructure delivers content using distinct microlevel strategies that are tied to that particular macrostructure type. Narratives use coherent and sequential text structures to present content. In contrast, expositories use text structures that pattern information according to the purpose of the expository text (e.g., sequential, cause and effect, compare and contrast). Because children have little experience with expository texts, it is likely that they will experience difficulty in learning from these programs.

Macrostructures organize content by pairing it with specific learning strategies to convey information. First, the effects of the two macrostructures on children's vocabulary and comprehension scores were evaluated for significant differences. Next, two sets of regressions were used to link strategies to outcomes for each macrostructure independently (i.e., comprehension strategies predicted outcomes for narratives and expositories separately). These analyses were performed to identify which strategies predicted what outcomes. All reported results are significant at alpha level $p < .05$ or better.

Method

Participants

After securing permission from principals at four elementary schools, teachers sent consent forms home with students of the target age. Parents of 71 second- and third-grade children attending four Title 1 public schools in two large Midwestern cities provided consent. Forty-two were girls ($M = 7.5$ years) and 29 were boys ($M = 7.8$ years). This age was recruited because the children had at least 1 year of basic reading instruction, resulting in some ability to decode words, a necessary skill for the original closed caption intervention (Linebarger, Piotrowski, & Greenwood, 2010). Eighty-two percent of the families reported incomes below $30,000, 5 children had an identified reading/learning disability, and 54.3% of the children had oral reading rates below a benchmark indicating risk for reading failure (Good & Kaminski, 2002; second-grade at-risk threshold ≤ 43 wpm, third-grade at-risk threshold ≤ 76 wpm). There were 52.6% at-risk second graders (i.e., at-risk $M = 22.3$ wpm, all other second graders $M = 71.2$ wpm) and 61.5% at-risk third graders (i.e., at-risk $M = 45.3$ wpm, all other third graders $M = 93.6$ wpm). Children were recruited from urban schools in predominantly low-socioeconomic-status (SES) communities: Fifty-six percent were African American children who spoke English as their first language and the remaining 44% were Hispanic children learning English as a second language (i.e., English language learners).

Procedure

After obtaining parental consent, all participating children completed a series of pretest assessments to measure baseline literacy skills. Following pretesting, children viewed one episode per visit across 12 visits. Video stimuli were presented to groups of 2 to 3 children using a 13-in. TV/VCR combination unit positioned approximately 1–2 ft away from them. Experimenters sat to the right and slightly behind the children and were directed to interact very little except to redirect attention to the screen in the event that it became necessary. The viewing took place in empty classrooms during the after school program hours at each of the schools. After viewing each episode, children participated in a short, individually administered posttest (i.e., there were enough research assistants to test all children immediately after viewing).

Measures

Program-specific vocabulary knowledge. Vocabulary knowledge was assessed in two ways: word fluency and definition knowledge. Children read five target words selected from each video after viewing the video. All words from each program were counted, categorized, and selected if they connected to an episode's main themes but were not function words (e.g., articles, prepositions, auxiliary verbs, participles; Watt, 2007). With these criteria, 60% of the words were nouns (e.g., *notch, warrior*), 27% were verbs (e.g., *arrive, locate*), and 13% were adjectives (e.g., *thirsty, rude*). Examiners scored children's word reading for accuracy and fluency (4-point scale: 0 = *inaccurate*, 1 = *read with difficulty*, 2 = *sounded out*, 3 = *fluent*; Word Fluency). Examiners were trained and practiced until interexaminer agreement on fluency scores equaled 90% or higher. After reading each target word, children were then asked to provide a definition. If the child was unable to read the target word, a clear pronunciation was provided prior to asking for the definition. Definitions were scored on a 3-point scale (0 = *no or inaccurate response*, 1 = *partially correct definition*, 2 = *accurate and complete definition*; Definition Knowledge). Two independent raters scored 11% of the definitions. Krippendorff's alpha indicated strong intercoder reliability for all definition scores ($\alpha = .93$).

Program comprehension. Comprehension questions included a literal and an inferential question for each episode. The literal question asked children to identify a critical story event that was visually and verbally presented (e.g., "When X happened what happened next?") while the inferential question required the children to provide the main idea of the program (i.e., "What do you think [program] was trying to teach you?"). Comprehension questions were scored on a 3-point scale (0 = *no/inaccurate response*, 1 = *partially correct answer*, 2 = *accurate/complete answer*). Two independent raters scored 11% of the answers. Krippendorff's alpha was acceptable for both literal comprehension ($\alpha = .91$) and inferential comprehension ($\alpha = .86$).

Baseline vocabulary knowledge and literacy ability. Prior to viewing any videos and after viewing all 12 videos, children read a list of all target words ($n = 60$ words). Although not the focus of this article, children demonstrated significant growth in video-specific word fluency scores from pre- to posttest (i.e., 16.5% growth; Linebarger et al., 2010). General literacy ability was measured using two DIBELS tasks: Nonsense Word Fluency (NWF) and Oral Reading Fluency (ORF; Good & Kaminski, 2002). Mean performance was 52.8 letter-sounds per minute (NWF; $SD = 29.3$) and 48.9 wpm (ORF; $SD = 32.1$).

Analytical Approach

Each validation outcome (i.e., Word Fluency, Definition Knowledge, Literal Comprehension, Inferential Comprehension) was z-score transformed to facilitate comparisons across literacy skills in each relevant analysis. Next, to control for potential third variables associated with the child's baseline literacy ability, a covariate was constructed by z-score transforming scores on the two pretest DIBELS literacy tasks and then summing these variables. Because these analyses were taken from a larger intervention study evaluating the effectiveness of closed captions on the literacy skills of two different populations (i.e., low-SES African American children and English language learners; Linebarger et al., 2010), we included two controls to extract the variance from the relations among strategies and child outcomes: intervention condition (i.e., one group saw all videos with captions while the other saw the same videos without captions) and the child's primary language. In preliminary analyses, the patterns of results were the same for both English- and Spanish-speaking children (i.e., all results were in the same direction; however, the findings were more robust for English language learners); therefore, adopting this strategy (i.e., controlling for primary language) appeared reasonable. Gender was unrelated to any of the outcomes and was dropped from further analyses.

To assess potential dependency issues resulting from viewing all 12 videos, two sets of intraclass correlations were calculated: between-subjects differences across children (i.e., .12–.16) or between-subjects differences across video (.92–.94). Larger intraclass correlations indicate that either videos or children (depending on the analyses performed) are more likely to share a common experience resulting in violation of the independence assumption. Given the larger variation between videos versus between children, dependency associated with individual performance across the 12 videos was minimal (Guo, 2005).

Results and Discussion

Two repeated measures analyses of covariance (ANCOVAs) were computed to examine child outcomes by macrostructure. Table 3 contains means, standard deviations, and ANCOVA results. Although there were no significant differences for literacy environment codes, it was possible that macrostructures would mediate the relations between codes and outcomes; therefore, regressions predicting outcomes using three sets of codes (i.e., literacy environment, comprehension strategies, vocabulary strategies) were conducted separately for each macrostructure. Both unstandardized (*B*) and standardized (β) coefficients from the regression models are found in Table 4. An alpha level of .05 was used for all statistical tests. Only interaction and main effects relevant to the hypotheses are presented next.

The Role of Macrostructure on Comprehension and Vocabulary Knowledge

Program comprehension. Based on Study 1 findings, we hypothesized that comprehension scores would be higher after viewing narratives compared with viewing expositories. To test this hypothesis, a 2 × 2 (Macrostructure × Comprehension Outcomes) repeated measures ANCOVA with repeated measures on the last factor was computed. Our hypothesis was supported. Comprehension was higher for narratives compared with expositories although the effect was qualified by a significant two-way

Table 3
Means, Standard Deviations, and Analysis of Covariance (ANCOVA) Results for Outcomes Alone and by Macrostructure

Outcomes	*M*	*SD*	Range
Word fluency	28.63	25.02	0–120
Definition knowledge	33.44	19.86	0–133
Literal comprehension	1.46	0.85	0–4
Inferential comprehension	1.69	0.68	0–4

	Narrative		Expository		ANCOVA (*F*-value)		
Outcomes	*M*	*SE*	*M*	*SE*	MS	O	MS × O
Vocabulary knowledge							
Word fluency	28.02	0.88	29.25	0.88	4.40* (0.005)	37.21*** (0.042)	21.41*** (0.025)
Definition knowledge	36.23	0.92	30.66	0.92			
Comprehension type							
Literal	1.64	0.04	1.29	0.04	11.77*** (0.014)	2.29 (0.003)	37.05*** (0.042)
Inferential	1.65	0.03	1.73	0.03			

Note. MS = Macrostructure; O = Outcome. Values in parentheses represent partial eta squared (η_p^2).
$^{*}p < .05$. $^{***}p < .001$.

Table 4
Regression Coefficients Predicting Outcomes Using Coded Strategies

	Narrative regression coefficients								Expository regression coefficients							
	WF		DK		LC		IC		WF		DK		LC		IC	
Literacy environment																
PU	−0.79***	(−0.26)	−0.93***	(−0.39)	0.03**	(0.23)	0.01	(0.05)	−0.15	(−0.05)	−0.81***	(−0.36)	0.03***	(0.35)	0.006†	(0.09)
PS	2.93**	(0.24)	3.35***	(0.36)	0.09*	(0.21)	0	(0.00)	−1.28*	(−0.09)	1.49**	(0.13)	−0.02	(−0.05)	−.004	(−0.01)
OP	1.35***	(0.34)	1.98***	(0.64)	0.05***	(0.33)	0.003	(0.02)	−0.32***	(−0.26)	−0.34***	(−0.34)	0.01**	(0.14)	−.006***	(−0.19)
R^2	0.56***		.25***		0.26***		0.01		0.49***		0.27***		0.13***		0.06***	
Comprehension strategies																
MS	−2.31**	(−0.17)	−3.48***	(−0.34)	−0.28***	(−0.58)	0.05	(0.11)	−2.53†	(−0.24)	11.90***	(1.32)	−0.43***	(−1.24)	0.03	(0.10)
PK	0.02***	(0.16)	0.03***	(0.31)	0.0002	(0.06)	0.001*	(0.15)	0.03*	(0.17)	0.22***	(1.44)	−0.006***	(−0.95)	0.0003	(0.08)
SM	−0.002	(−0.03)	−0.002	(−0.05)	−0.0002†	(−0.10)	−0.0***	(−0.23)	0.01	(0.18)	−0.07***	(−1.49)	0.002***	(1.19)	0.00001	(0.01)
QP	0.65***	(0.32)	0.97***	(0.62)	0.03***	(0.38)	−0.01*	(−0.22)	0.25	(0.13)	−1.09***	(−0.69)	0.04***	(0.65)	−0.01†	(−0.22)
R^2	0.57***		0.27***		0.24***		0.05**		0.52***		0.49***		0.18***		0.05**	
Vocabulary strategies																
TW	0.19	(0.07)	0.36	(0.16)	−0.07***	(−0.69)	0.02†	(0.25)	0.41	(0.10)	−3.64***	(−1.06)	0.12***	(0.87)	−0.02	(−0.15)
NR	0.24	(0.07)	−0.09	(−0.03)	−0.08***	(−0.58)	−0.03	(−0.23)	−0.53**	(−0.29)	1.05***	(0.71)	−0.05***	(−0.89)	0.006	(0.13)
SL	−1.25	(−0.04)	−6.49	(−0.25)	0.25	(0.20)	−0.94**	(−0.88)	−12.98***	(−0.49)	10.37***	(0.49)	−0.59***	(−0.71)	0.22†	(0.34)
%D	−1.76	(−0.18)	−0.49	(−0.07)	−0.15	(−0.42)	0.37***	(1.21)	−4.21***	(−0.53)	2.07**	(0.33)	−0.13***	(−0.54)	0.01	(0.07)
R^2	0.56***		0.28***		0.28***		0.06***		0.53***		0.48***		0.20***		0.06***	

Note. Both unstandardized (*B*) and standardized (β) regression coefficients are presented. β coefficients are displayed in parentheses.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

(Macrostructure × Comprehension Outcomes) interaction. Literal comprehension was higher after viewing narratives while inferential comprehension was higher after viewing expositories. Univariate tests indicated that differences were significant for literal comprehension, $F(1, 847) = 36.70$, $\eta_p^2 = .042$, and marginally significant ($p < .10$) for inferential comprehension, $F(1, 847) = 2.85$, $\eta_p^2 = .003$.

Program-specific vocabulary knowledge. Based on Study 1 findings, we hypothesized that program-specific vocabulary knowledge would be higher after viewing expositories compared with viewing narratives. To test this hypothesis, a 2 × 2 (Macrostructure × Vocabulary Outcomes) repeated measures ANCOVA with repeated measures on the last factor was conducted. Our hypothesis was not supported. Main effects for Vocabulary Outcomes and Macrostructure were significant. Definition knowledge was higher than word fluency while vocabulary knowledge overall was higher after viewing narratives compared with expositories.

As with comprehension, the Macrostructure main effect was qualified by a significant two-way (Macrostructure × Vocabulary Outcomes) interaction. As predicted, word fluency was higher after viewing expositories. Contrary to our prediction, definition knowledge was higher after viewing narratives. Univariate tests indicated that these differences were significant for definition knowledge, $F(1, 847) = 18.52$, $\eta_p^2 = .021$, and not significant for word fluency, $F(1, 847) = 0.98$, $\eta_p^2 = .001$.

What Is the Relation Between the Literacy Environment and Children's Program Comprehension and Program-Specific Vocabulary Knowledge by Macrostructure?

Recall that character's use of and positive comments about print and onscreen print were hypothesized to create a conducive learning environment. For narratives, two of the three codes predicted outcomes. Positive statements about print and onscreen print positively predicted literal comprehension and both vocabulary outcomes while characters' print use supported literal comprehension and suppressed word fluency and definition knowledge.

Unlike narratives, there were no consistent patterns between the literacy environment and outcomes within expositories. Each time characters used print, definition knowledge scores dropped, literal comprehension improved, and inferential comprehension marginally improved. Positive statements about print supported definition knowledge and suppressed word fluency. Onscreen print negatively predicted definition knowledge, word fluency, and inferential comprehension and positively predicted literal comprehension.

What Is the Relation Between Comprehension Strategies and Children's Comprehension and Vocabulary Knowledge by Macrostructure?

There were four comprehension strategies: main segments, time spent directing viewers to access prior knowledge, time spent summarizing a main segment, and total questions/predictions used to highlight upcoming content. Results were fairly consistent for narratives; that is, most coefficients were in the same direction (although some were unrelated) across outcomes. As main segments increased, literal comprehension and both vocabulary scores decreased. Spending time activating prior knowledge predicted higher vocabulary and inferential comprehension while time spent summarizing main content suppressed literal (marginally) and inferential comprehension. As characters' questions and predictions increased, vocabulary performance and literal comprehension increased while inferential comprehension decreased.

For expositories, the results were less consistent. As main segments increased, word fluency and literal comprehension suffered while definition knowledge improved. Activating prior knowledge both positively predicted vocabulary outcomes and negatively predicted literal comprehension. In contrast, longer summaries predicted lower definition knowledge and higher literal comprehension. As the number of questions and predictions increased, definition knowledge and inferential comprehension decreased and literal comprehension increased.

What Is the Relation Between Vocabulary Strategies and Children's Program Comprehension and Program-Specific Vocabulary Knowledge by Macrostructure?

Vocabulary strategies consisted of four coded variables: frequency of target word use, average sentence length, percentage of Dolch words, and the narration rate. Interestingly, for narratives, none of the vocabulary strategies predicted word fluency or definition knowledge. Literal comprehension declined as repetition of target words and the narration rate increased. Inferential comprehension declined as sentence length increased while it improved as the percentage of Dolch words increased.

For expositories, the role of vocabulary strategies was much more prominent. Word fluency scores decreased as the narration rate, the sentence length and the percentage of Dolch words increased. Definition knowledge scores decreased with each repetition of a target word and increased with faster narration rates, longer sentences, and a higher percentage of Dolch words. Literal comprehension increased with each repetition of a target word, and decreased with faster narration rates, longer sentences, and a smaller percentage of Dolch words. Inferential comprehension scores marginally increased with longer sentences.

General Discussion

In Study 1, a set of learning strategies were identified and coded across two episodes of six children's TV programs. Programs were further categorized using one of two macrostructures: narrative or expository. In Study 2, working with a population of children from low-SES homes at risk for literacy difficulties, we evaluated the general effects associated with watching content presented within each macrostructure as well as whether and how the coded strategies embedded in each macrostructure predicted vocabulary and comprehension.

The Role of Narrative and Expository Macrostructures

Study 1 established that narrative macrostructures contained more learning strategies hypothesized to support comprehension outcomes while expository macrostructures contained more of both supportive and unsupportive strategies hypothesized to predict vocabulary outcomes. We further hypothesized that narratives would predict stronger comprehension performance while expositories would predict stronger vocabulary performance. Study 2 partially confirmed these hypotheses. When viewing narrative macrostructures, all outcomes were higher compared with outcomes when viewing expository macrostructures.

Programs using narrative macrostructures tell stories with a beginning, middle, and end; use characters; include plot or conflict; and create a setting. In other words, they present content in a structurally homogenous way using prototypical story parts. Narrative macrostructures should reduce processing demands due to children's considerable narrative experience in real-life, print-based, and TV-based contexts (Fisch, 2004; Linebarger & Piotrowski, 2009). This narrative experience likely translated into both sensitivity to and automaticity with the structures typical of narratives, making children faster and more competent at identifying story events (i.e., literal comprehension), filling in gaps in presented content, and freeing up valuable cognitive resources to make connections to real-life situations or broader themes (i.e., inferential comprehension; Smith & Ragan, 2005). As predicted, identification of literal story events was stronger for narrative macrostructures. Although narratives were expected to make it easier to refocus cognitive resources toward making broader connections, children infrequently did this (Linebarger, 2001). Previously, deep processing was more likely to occur if content was moderately novel and of intermediate complexity while superficial processing occurred with content that was overly familiar or excessively challenging (Rice et al., 1982; Roller, 1990). With experience, processing narratives becomes automatic and children are more likely to allocate minimal resources to them.

Unlike the prototypical format of narratives, programs that use an expository macrostructure present content in structurally heterogeneous ways (Best et al., 2008). Based on Study 1, watching expositories was expected to predict stronger vocabulary outcomes when compared with narratives because expositories used more supportive vocabulary strategies. Contrary to this prediction, vocabulary performance (particularly definition knowledge) was higher after viewing narratives versus expositories. The structural heterogeneity inherent in expositories offers an explanation. When processing unfamiliar content delivered through an expository macrostructure, children must simultaneously process the new content and the general purpose of an expository text to effectively learn the content. Duke and Kays (1998) contend that young children can learn from expository texts but, because they are not consistently exposed to these formats until as late as fourth grade, learning from expository texts is difficult and frustrating.

The Role of Learning Strategies

In Study 1, three categories of learning strategies embedded in children's educational TV were identified and coded: strategies depicting a supportive literacy environment, strategies supporting comprehension, and strategies supporting vocabulary. Despite selecting strategies with evidence of their effectiveness in other contexts, the hypothesized effects were not uniformly present. Instead, relations between strategy and outcome differed by strategy type, program macrostructure, specific child outcome, or some combination of these three factors. Refocusing interpretation to the cognitive activity generated by a strategy clarified the inconsistent findings. Smith and Ragan (2005) proposed five broad types of learning strategies: organizing, rehearsing, elaborating, metacognitive, and affective. The first four strategies elicit specific cognitive activity while affective strategies create environments that trigger attention to and increased engagement with content. No metacognitive strategies were coded in this study.

An additional layer of complexity was introduced by the macrostructures used to deliver content. As described, macrostructures introduce a framework for content through their use of particular and predictable learning strategies that arrange ideas and make connections among these ideas. Narrative macrostructures use a fairly uniform set of strategies and children's accumulated experiences with narratives are substantial. Coupling predictability with experience suggests that narratives may be better macrostructures for delivering content, especially when that content is unfamiliar or complex. Unlike narratives, expository macrostructures are highly variable and dependent on the purpose of a program or the unique characteristics of particular content. Expositories can include strategies that describe a topic, present sequential or chronological information, introduce cause and effect, or highlight similarities and differences. Because young children have little experience with expository forms, learning content embedded with these forms may prove challenging. Recent evidence suggests that introducing expository text in preschool or early elementary school can help children become familiar with a range of expository forms and subsequently mitigate difficulties that arise when children shift from learning to read to reading to learn around fourth grade (Duke & Kays, 1998).

Organizing strategies structured content by encasing it within a common framework and by using simple supporting text to increase attention to conceptual content (Smith & Ragan, 2005). In this study, simple supporting text was defined by more high-frequency words (i.e., Dolch words), longer but less complex sentences (sentence length), and narration rates that did not exceed the average wpm that children this age are able to adequately process (i.e., narration rate; Barron, 2004). These strategies worked in concert to increase or decrease the volume of content contained in an episode. Greater use of main segments was associated with more unique content ideas that structured the presentation and flow of that content. As main segments increased, the narration rate sped up considerably while the sentence length modestly increased. Sentence length was more highly related to Dolch words, suggesting that sentences became longer by using more high-frequency words while narration rates became faster due to more content words. The latter relation indicates cognitive overload.

Narratives with more main segments negatively predicted word fluency, definition knowledge, and literal comprehension likely because extra main segments were indicative of more irrelevant subplots, making it hard to keep the plot points together. Both vocabulary outcomes and literal comprehension suffered as a result. These difficulties were compounded when dialogue was spoken faster and sentence lengths were longer. Simplifying the content came at a cost as definition knowledge suffered. Without key conceptual content or extended discussions about particular concepts, viewers were able to provide limited definitions. Inference generation was supported with shorter dialog segments filled with more high-frequency words.

Expositories contained more main segments, longer sentences, more Dolch words, and faster narration rates. As a result, word fluency and literal comprehension suffered. In contrast, the conceptual understanding needed to provide solid word definitions was supported by more volume. Main segments tended to build on each other throughout an expository episode. The three expositories used in this study contained multiple examples and descriptions that linked to the theme of that episode. For example, in RR, one episode about humpback whales featured multiple vignettes that defined this type of whale from several viewpoints, including a scientist, a whale-watching boat captain, the host, and a nonfiction book.

Rehearsing strategies directly supported the encoding and retrieval of information by providing multiple opportunities for practice. Coded strategies were repetition of target words and repetition of main segment content (i.e., summaries). When content is repeated, it signals that it is important and should be attended to. Repeating target words is a fairly simple rehearsal tool while providing main segment summaries is a more complex rehearsal strategy.

When narratives used more rehearsal strategies, comprehension was better. Children's experience and facility processing narratives freed up resources to devote to comprehending content. By spending time repeating content, children were better able to identify story events (i.e., literal comprehension) and, more important, generate inferences about these events.

Rehearsal strategies embedded in expository macrostructures aided acquisition of definition knowledge and, to a lesser extent, literal comprehension. Expositories contained noticeably more content than narratives. To ensure adequate comprehension of vocabulary, expositories included multiple rehearsal opportunities (i.e., words averaged 27 repeats and summaries averaged 8 min). Interestingly, summaries were almost 2 min longer in narratives, a finding that may explain why outcomes were higher after viewing narratives.

Elaborating strategies help learners create links between new content and existing prior knowledge. Two strategies elicited this cognitive activity: time spent activating viewers' prior knowledge and total questions and predictions made about upcoming content. Both strategies enabled viewers to establish associations between new information and prior knowledge, to identify key concepts or story events, and to increase the personal relevance of the content.

When narratives used more elaboration, all outcomes were higher especially vocabulary. As with the other cognitive strategy types, when used in a narrative macrostructure, children were able to devote their attention to conceptual information and story events rather than needing to split time simultaneously to make sense of the presentation format and the content.

Elaboration strategies used in expositories predicted more definition knowledge and inferential comprehension. Elaborating is a demanding cognitive activity because it requires that children understand program content, possess knowledge about how content links to the macrostructure, and connect content and strategy to relevant background knowledge. Using too many elaborating strategies impeded literal comprehension likely because they directed attention to conceptual understanding and away from key events. In the comprehension hierarchy, knowing story events is less important over time while generating inferences or themes is more important.

Frequently using affective strategies should influence engagement with and motivation toward attending to content resulting in a more favorable learning environment. Three affective strategies were hypothesized to create such an environment: characters using print, characters making positive statements about print, and the display of onscreen print. These strategies were conceptually linked to three instructional conditions found to influence learning: demonstration of the desired behaviors by a respected role model, opportunities for practice using those behaviors, and reinforcement of those behaviors (Gagne, 1985; Smith & Ragan, 2005).

Narratives supported program-specific learning (i.e., word fluency, definition knowledge, literal comprehension) when characters made positive statements about print or text was displayed onscreen. Characters using print hindered literal comprehension and both vocabulary outcomes. The characters using print may not be likeable or credible or there may have been too little explicit reinforcement of the benefits of print (Gagne, 1985; Smith & Ragan, 2005). These explanations are supported by the positive relation between positive statements and outcomes. Onscreen print provided opportunities for reading practice and elaboration by providing visual and aural cues that make print more salient and easier to decode. While the literacy environment helped children learn program-specific content, this learning did not transfer to inferencing.

Connections between affective strategies and outcomes were less consistent for expositories. Modeling print use predicted both comprehension outcomes but impeded definition knowledge while positive statements predicted stronger definition knowledge and impeded word fluency. Onscreen print supported literal comprehension only. The remaining three outcomes were diminished by more onscreen print. These inconsistencies are likely due to a combination of cognitive overload and gaps in knowledge brought on by unfamiliar forms and content.

Conclusion

The results of the two studies indicate that macrostructures and learning strategies must be considered simultaneously to understand how children learn from televised content. To facilitate learning, programs should generate cognitive activity that takes advantage of the broader presentation framework along with particular strategies that lead to learning when used in this framework. Over time and with experience, processing demands shift due to stimulus-specific and person-specific factors. Stimulus-specific factors included macrostructures and learning strategies used to convey content and generate cognitive activity. In contrast, person-specific factors included experience with macrostructures (particularly for expository texts) and accumulated background knowledge about the content and its presentation form. While children typically have more accumulated experiences with narratives, they can, when taught to do so, learn equally well from expository macrostructures. TV provides a unique opportunity to familiarize children with expository structures. There is clear evidence that TV comprehension is linked to reading comprehension; that is, TV comprehension was highly correlated with print comprehension in preschool (Linebarger & Piotrowski, 2009) and predictive of conventional reading success longitudinally (i.e., TV comprehension at 6 years predicted reading comprehension at 8 years; Kendeou et al., 2005). General processing skills used across media share many commonalities. As such, educational TV can play a vital role in preparing at-risk children in low-SES families to read.

References

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior: The Recontact Study. *Monographs of the Society for Research in Child Development*, *66*(1), 1–143.

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, *3*, 265–299.

Barron, A. E. (2004). Auditory instruction. In D. H. Jonassen (Ed.), *Handbook of research in educational communications and technology* (pp. 949–978). Mahwah, NJ: Erlbaum.

Best, R., Floyd, R., & McNamara, D. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, *29*, 137–164.

Crawley, A. M., Anderson, D. R., Wilder, A., Williams, M., & Santomero, A. (1999). Effects of repeated exposure to a single episode of the television program *Blue's Clues* on the viewing behaviors and comprehension of preschool children. *Journal of Educational Psychology*, *91*, 630–637.

Duke, N. K., & Kays, J. (1998). "Can I say 'once upon a time?'": Kindergarten children developing knowledge of information book language. *Early Childhood Research Quarterly*, *13*, 295–318.

Fisch, S. M. (2004). *Children's learning from educational television*. Mahwah, NJ: Erlbaum.

Gagne, R. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart & Winston.

Good, R. H., & Kaminski, R. A. (2002) *Dynamic indicators of basic early literacy skills (DIBELS)*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved February 11, 2008, from http://dibels.uoregon.edu/

Guo, S. (2005). Analyzing grouped data with hierarchical linear modeling. *Children and Youth Services Review*, *27*, 637–652.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77–89.

Herrman, D. J., Yoder, C., Gruneberg, M., & Payne, D. (2006). *Applied cognitive psychology: A textbook*. Mahwah, NJ: Erlbaum.

Jensema, C. J., McCann, R., & Ramsey, S. (1996). Closed-captioned television presentation speed and vocabulary. *American Annals of the Deaf*, *141*, 284–292.

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, *33*, 91–98.

Linebarger, D. L. (2001). Learning to read from television: The effects of using captions and narration. *Journal of Educational Psychology*, *93*, 288–298.

Linebarger, D. L., Kosanic, A. Z., Greenwood, C. R., & Doku, N. S. (2004). Effects of viewing the television program *Between the Lions* on the emergent literacy skills of young children. *Journal of Educational Psychology*, *96*, 297–308.

Linebarger, D. L., & Piotrowski, J. T. (2009). TV as Storyteller: How exposure to television narratives impacts at-risk preschoolers' story knowledge and narrative skills. *British Journal of Developmental Psychology*, *27*, 47–69.

Linebarger, D. L., Piotrowski, J. T., & Greenwood, C. R. (2010). Onscreen print: The role of closed captions as a supplemental literacy tool. *Journal of Research in Reading*, *33*(2), 148–167.

Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and relations to language outcomes. *American Behavioral Scientist*, *46*, 624–645.

May, F. (1998). *Reading as communication: To help children write and read*. Upper Saddle River, NJ: Merrill.

McMahon, M. L. (1983). Development of reading-while-listening skills in the primary grades. *Reading Research Quarterly*, *19*, 38–52.

Michel, E., & Roebers, C. M. (2008). Children's knowledge acquisition through film: Influence of programme characteristics. *Applied Cognitive Psychology*, *22*, 1228–1244.

Naigles, L., & Mayeux, L. (2001). Television as incidental language teacher. In D. G. Singer & J. L. Singer (Eds.), *Handbook for children and the media* (pp. 135–152). London: Sage.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-476). Washington, DC: U.S. Government Printing Office.

Nielsen Media Research. (2003). *2002/03 kids programming AA kids 2-5 rating ranker*. New York: Nielsen Media Research.

Nord, C. W., Lennon, J., Liu, B., & Chandler, K. (1999) *Home literacy activities and signs of children's emerging literacy, 1993 and 1999*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. Retrieved June 28, 2007, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000026

Patterson, J. L. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics*, *23*, 493–508.

Reinking, D., & Schreiner, R. (1985). The effects of computer-mediated text on measures of reading comprehension and reading behavior. *Reading Research Quarterly*, *20*, 536–552.

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 361–395). Hillsdale, NJ: Erlbaum.

Rice, M. L., Huston, A. C., & Wright, J. C. (1982). The forms of television: Effects on children's attention, comprehension, and social behavior. In D. Pearl, L. Bouthilet, & J. B. Lazar (Eds.), *Television and behavior: Ten years of scientific progress and implications for the 80's* (Vol. 2, pp. 24–38). Washington, DC: Government Printing Office.

Rice, M. L., & Woodsmall, L. (1988). Lessons from television: Children's word learning when viewing. *Child Development*, *59*, 420–429.

Roller, C. M. (1990). The interaction between knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly*, *25*, 79–89.

Sadoski, M., & Paivio, A. (2001). Imagery and text: A dual coding theory of reading and writing. *Reading and Writing*, *16*, 259–262.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, *14*, 245–302.

Schiefele, U. (1998). Individual interest and learning—What we know and what we don't know. In L. Hoffman, A. Krapp, & K. A. Renninger (Eds.), *Proceedings of the Seeon Conference on Interest and Gender* (pp. 91–104). Kiel, Germany: IPN.

Smith, P., & Ragan, T. (2005). *Instructional design* (3rd ed.). Hoboken, NJ: Wiley Jossey-Bass.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.

Stein, N. L., & Trabasso, T. (1982). What's in a story? An approach to comprehension. In R. Glaser (Ed.), *Advances in the psychology of instruction* (pp. 213–268). Hillsdale, NJ: Erlbaum.

Szabo, M., & Lamielle-Landy, A. (1981). Television-based reading instruction, reading achievement, and task involvement. *Journal of Educational Research*, *74*, 239–244.

Taylor, S. E. (1964). *Listening: What research says to the teacher*. Washington, DC: National Education Association of the United States.

Taylor, B. M., Peterson, D. S., Pearson, P. D., & Rodriguez, M. C. (2002). Looking inside classrooms: Reflecting on the ''how'' as well as the ''what'' in effective reading instruction. *The Reading Teacher*, *56*, 270–279.

Uchikoshi, Y. (2005). Narrative development in bilingual kindergarteners: Can *Arthur* help? *Developmental Psychology*, *41*, 464–478.

Watt, R. J. C. (2007). *Concordance software*. Retrieved August 28, 2007, from http://www.concordancesoftware.co.uk/

Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Pinon, M., Scantlin, R., et al. (2001). The relations of early television viewing to school readiness and vocabulary of children from low income families: The early window project. *Child Development*, *72*, 1347–1366.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1598–1619

Home Media and Children's Achievement and Behavior

Sandra L. Hofferth
University of Maryland

This study provides a national picture of the time American 6- to 12-year-olds spent playing video games, using the computer, and watching TV at home in 1997 and 2003, and the association of early use with their achievement and behavior as adolescents. Girls benefited from computer use more than boys, and Black children benefited more than White children. Greater computer use in middle childhood was associated with increased achievement for White and Black girls, and for Black but not White boys. Increased video game play was associated with an improved ability to solve applied problems for Black girls but lower verbal achievement for all girls. For boys, increased video game play was linked to increased aggressive behavior problems.

Recent reports emphasize the pervasive influence of the media on children's lives (Roberts, Foehr, & Rideout, 2005). In 1950, only about 9% of American families owned a TV set, but by the early 1980s, when the first systematic examinations of children's time began, 98% of American families owned one (Andreasen, 1994). Since then, there has been a dramatic increase in communication technologies available to the public, all of which are subsumed under the term *media*. Although they may not be new to many adults—three fourths of children lived in a home with a computer in 1999—the usage by children of home computers and video games, what are called *new media*, is relatively recent. In 1999, fewer than half (47%) of children 8–18 used a computer during the previous day, rising to 54% by 2004 (Roberts et al., 2005). In 1999, children 8–18 spent only 27 min in an average day using the computer for recreation at home; that had more than doubled to just over 1 hr by 2004. Another major change was in access to the Internet; the proportion of children living in a home with Internet access rose from 47% to 74% between 1999 and 2004 (Roberts et al., 2005).

As a result of the rapid diffusion of new technologies, research on their effects has been slow to catch up. A recent critique of a leading textbook on children and the media pointed to the lack of systematic information over time on children's use of TV, computers, and other media (Sotirovic, 2005). Comparing children's media use over time is difficult because research reports do not use comparable samples of children and often do not break out results for age groups. In addition, they permit overlap in use. Not adjusting media time for simultaneous use leads to exaggerated estimates of total time spent with media, implying a day of more than 24 hr. Most studies fail to include reliable measures of what else children are doing and what trade-offs between activities are made. Finally, few studies focus on middle childhood, an important period in establishing long-term media habits, one in which large changes in media use are occurring, and one in which the family context is critical.

The present research is the first to examine the implications for children of increased exposure to media at home using nationally representative data. It reports changes in the past decade in the utilization of computers, video games, and TV in comparison with other activities among children in middle childhood. Second, it examines how the utilization of new media is tied to other aspects of children's lives. The introduction of new technologies may change children's TV viewing, alter study habits, and change leisure time pursuits and sleep patterns. Using a widely accepted methodology that ensures that total time not exceed 24 hr in a day and that obtains information not only on media but also on other activities, it examines the association of new media time with playing, studying, reading,

---

Support for this study was provided under NICHD Grants R21-HD-050125 and R24-HD041041.

Correspondence concerning to this article should be addressed to Sandra L. Hofferth, Department of Family Science, School of Public Health, University of Maryland, College Park, MD 20742. Electronic mail may be sent to hofferth@umd.edu.

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0020

sleeping, sports, TV viewing, and outdoor time. Third, this research examines the consequences of new media use at ages 6–12 for children's achievement and behavior in adolescence 6 years later.

Theory and Review of Literature

Theory About the Effects of Media on Children

The introduction of any new form of media technology leads to a debate about its implications, especially if children have access. Critics of computer use by children focus on two concerns. First, computer use per se may not contribute to development because children learn through creative and imaginative play, which require manipulating the physical environment (Cordes & Miller, 2000; Healy, 1998). Second, computers may displace traditional learning-related activities, such as reading, studying, and nonscreen play that are essential for healthy child development (Van Evra, 2004). However, others argue that media may benefit children. Computer literacy has been touted as the primary avenue to success in the 21st century (Attewell, 2001). Learning how to use computers has potential advantages; most adult jobs require at least minimal computer literacy and many jobs require more advanced computer skills. Playing educational games and searching for information on the Internet can improve children's academic and computer skills.

Theory provides several avenues for potential consequences of media exposure. The first pathway is *social or observational learning* (Bandura, 2002). Media can add to development if they provide new information or stimulate children's learning. Children have been found to gain knowledge about the world and to learn directly and indirectly from media about how to solve problems (Fisch, 2004; Van Evra, 2004). Research using brain imaging documents areas of the brain that are activated in viewing TV, areas associated with memory, face and object recognition, attention, intention, perception, and arousal in threat situations (Anderson, Fite, Petrovich, & Hirsch, 2006). Children also learn how to find information and to manipulate media devices, skills that transcend content.

Research has failed to prove a strong link between computer use and child learning. One study found that children with a home computer reported higher grades than those without a home computer (Rocheleau, 1995); however, other studies found no effects of home computer use on achievement (Subrahmanyam, Greenfield, Kraut, & Gross, 2001). There is one exception; drill and practice programs for learning specific skills have been linked to the skills targeted (Coley, Cradler, & Engle, 1997). Unfortunately, there are only a few studies and the outcomes are mostly self-reported. Another major limitation is that the research to date has failed to specify the use of the computer—for playing games, studying, or communication.

Research has shown that educational TV viewing boosts early literacy and reading habits (Anderson, Huston, Schmitt, Linebarger, & Wright, 2001; Fisch & Truglio, 2001). Besides the benefits that have been found for very young children, educational programs have also been shown to have raised motivation to read, increased knowledge of current events, and improved problem-solving skills in math and science among older children (Fisch, 2004). Given the record of learning from educational TV programming and given the highly interactive nature of games, educational computer or video games have the same or greater potential for learning. Game playing has potentially positive effects, such as improved motor skills, spatial skills, iconic skills (the ability to read images), practice in logic, problem solving, visual attention, and active learning, practice, and discovery leading to mastery (Gee, 2003; Van Evra, 2004). Research has corroborated the effectiveness of game playing for improving spatial skills and visual-perceptual skills, such as improving hand–eye coordination (Subrahmanyam & Greenfield, 1994). Although the emphasis has been on visual and spatial skills, children may seek and read online tips for improving their game strategies and skills. Many games require the same skills that are tested in nonverbal IQ tests (Subrahmanyam et al., 2001) plus a high level of involvement and concentration (Lieberman, 2006).

A second theory, *cultivation theory*, argues that media content affects viewers' beliefs about the world and, consequently, alters their behavior (Gerbner, Gross, Morgan, Signorielli, & Shanahan, 2002). For example, exposure to violent media leads to a belief that aggression can resolve problems with no adverse effects. Under this theory, more time exposed to violent games would lead to more aggressive behavior (Anderson et al., 2001). The gender-stereotyped characteristics and behaviors of game characters reinforce traditional notions of male and female roles (Alloway & Gillbert, 1998; Gerbner et al., 2002), which may lead to lower self-esteem among female players (Lee & Peng, 2006).

Although educational, fantasy, and sports games exist, even those with the broadest age rating can contain mild violence. The pervasiveness of violent and aggressive video games suggests that more exposure means more exposure to violence. A meta-analysis of 35 research reports concluded that exposure to violent games increased later aggressive behavior for both boys and girls (Anderson & Bushman, 2001). Violent game play was associated with aggressive thoughts, aggressive affect, and physiological arousal, with suppression of affective areas of the brain responsible for positive emotions (Weber, Ritterfeld, & Mathiak, 2006). Exposure was more harmful for youth who were more aggressive to begin with (Anderson & Bushman, 2001; Van Evra, 2004). If computers are used to play games, the same outcomes are possible.

Use and motivation theory argues that the impact of media depends upon the uses to which media are put, the motivations (learning, entertainment), and the individual abilities and dispositions of the user (Rubin, 2002). Thus, the context and selectivity of media use need to be taken into account. Young children are more susceptible to influence because their attitudes, beliefs, ideas, and knowledge are developing rapidly and they absorb information from every source. They cannot be expected to have the same sophistication of understanding as do older children or adults. Toddlers, for example, have been shown to learn less quickly and consistently from behaviors modeled on a TV screen than those in real life (Schmitt & Anderson, 2002). Computer use shifts with age; younger children play games whereas older children spend more time accessing Web sites and communicating by e-mail and instant messages (Roberts et al., 2005). Similarly, video game play changes with age. Educational games decline and sensorimotor games increase (Wright et al., 2001). Beginning in about fifth grade, the motivation for video game playing is challenge and fantasy. As children age into junior high school, social motivation comes into play, including interaction around games, competition, and sharing of strategies (Van Evra, 2004).

Media use is highly gendered. Although there are few differences in TV viewing or computer use, boys devote more time to video game play than girls. An important aspect of game playing, both video game and computer game play, is the reinforcement of gender-differentiated peer group culture (Sefton-Green, 1998). Games may provide opportunities to compete and cooperate around game playing, thus facilitating social interaction. Male talk often centers on discussions of games and problem solving (Vered, 1998). Because in most games the main character is male and the activities are also male oriented (von Salisch, Oppl, & Kristen, 2006), it should not be surprising that boys develop a greater interest in and expertise in video games and spend more time than girls playing such games (Roberts et al., 2005). Boys prefer more aggressive games whereas girls prefer well-developed characters and stories and bring different expectations and norms to the play setting, eschewing yelling and aggression (Vered, 1998). Because of the different amount and type of video game play by boys and girls, their behaviors may be differentially influenced (Subrahmanyam & Greenfield, 1994). Greater use of educational game play may give girls' achievement a boost (von Salisch et al., 2006). For boys, game playing could lead to involvement in supportive peer groups rather than to isolation. This may be especially the case for video games, which provide multiple controllers, compared with computer games that have only one mouse. For girls, game playing could lead to greater social isolation, especially if they play games on the computer rather than on hand-held consoles. For both genders, greater use of the computer for communications, such as e-mail and instant messaging may lead to greater social involvement.

Research shows African Americans and Hispanics to have greater overall media exposure; however, this is due to the latter's heavier time spent watching TV. Substantial racial/ethnic differences in home computer access have been documented, with African American and Hispanic 2- to 7-year-olds averaging only 60% of the usage of White children of the same age (Roberts, Foehr, Rideout, & Brodie, 1999). Because of their lower access to other learning-related resources, children from minority families may benefit more from having access to a computer at home than children from majority families (Wright et al., 2001). As children enter school, learning resources increase, but most time is still spent at home. Black children are also disadvantaged at school; children living in zip codes with an annual median income below $25,000, more likely to be the residential location of Black and Hispanic families, had about one fourth of the exposure to computers of those with incomes over $40,000 (Roberts et al., 1999). African American boys may not experience the same peer group culture that focuses on video games as do White boys. Because most of the early media research was conducted using White middle-class samples in the Midwest, we have less information about differential effects of media use among different race/ethnic groups. Thus, it is important to investigate whether there are differential effects of computer and video game use across racial/ethnic groups separate from income. That is, the effect of media use on reading and math scores and behavior may vary by race, independent of income and education (Attewell & Battle, 1999).

Parental characteristics and behavior may moderate the effects of media (Nikken & Jansz, 2006) because children are dependent upon them for both media and control over their subsequent behavior. Children whose parents have less education and whose families have lower incomes have been shown to have less access to computers than their more advantaged peers (Attewell, Suazo-Garcia, & Battle, 2003; Roberts et al., 1999; Roberts et al., 2005).

In sum, greater exposure to media facilitates learning the what (content) and how (technical skills) of the information displayed. It also promotes socialization into peer cultures. Media's effects are dependent upon the context in which their use takes place and characteristics of child and family (e.g., age, gender, and race; Fisch, 2004). In addition, whether media are used for game play, study, or for communication may influence outcomes.

Consequences can also occur because media used for entertainment or social activities may displace activities such as studying, reading, physically active sports and outdoor activities, nonscreen play, and sleep, which positively contribute to health and development. Thus, the fourth theory, *displacement theory*, is that media use displaces important developmental activities (Anderson et al., 2001; Neuman, 1995). A negative association is expected; because time is limited, if you do more of one type of activity you have less time for other activities. *Which activities are displaced and by how much is the important issue.* There are at least three possible mechanisms of displacement: First, new media displace functionally equivalent activities (e.g., video games displace computer games); second, new media replace similar but less structured activities (e.g., computer play displaces nonscreen play or outdoor activities); and third, new media displace less satisfying activities or activities with smaller benefits at the margin (e.g., computer games displace studying or sleeping).

One of the major concerns of developmental psychologists is that the use of computers and video games could reduce children's social involvement and increase isolation. The limited research available has been unable to document a negative association of computer use with social behavior, perhaps because some social interactions, such as communication with friends, are facilitated by computers (Subrahmanyam et al., 2001). A recent study that provided home computers and Internet access to families who did not previously have such access found significant short-term (1-year) declines in social involvement but these effects weakened after the 1st year (Subrahmanyam et al., 2001). Research has found that increased computer and video game use altered interaction patterns as families incorporated media into their lives (Mitchell, 1985; Nikken & Jansz, 2006), but there was no overall negative impact. Several studies found greater computer use by children associated with watching less TV (Neuman, 1995; Wright et al., 2001) and another found greater computer use associated with less time in sports and outdoor activities (Attewell et al., 2003).

Small amounts of media use are unlikely to indirectly affect achievement and behavior because they do not displace much other activity time. Thus, a curvilinear effects hypothesis is reasonable. That is, although low to moderate usage could promote positive achievement and behavior through social learning, *heavy usage* could displace time spent studying, reading, playing, and interacting with peers, and this may lead to reduced achievement and increased social isolation, low self-esteem, and inability to relate to others (Lee & Peng, 2006; Subrahmanyam et al., 2001). One study found low (but not high) computer use to be associated with more time spent reading and greater letter–word recognition, reading comprehension, and mathematical calculation achievement test scores among elementary schoolchildren (Attewell et al., 2003). However, another study found heavy computer use to be associated with greater achievement by adolescents (Rocheleau, 1995).

Finally, video game play may displace the typical activities in which children engage, leading to social isolation or overdependence on games. More time in video games is likely to interfere both with more accepted learning-related activities and with physical and social activities, such as sports and informal outdoor activities. One study found video game play leading to less prosocial behavior, such as donating to others (Chambers & Ascione, 1987). Because no prior research has examined the displacement effect and there is little known about its association with achievement, more systematic research is needed.

Are Media Effects Causal?

Does media use reduce participation in other activities, does having fewer other activities lead to increased media use, or does something else both reduce other activities and increase media use? Children and their parents choose how to spend their time. TV and other media serve as entertainment, a way for families to spend time together, and as a source of educational, social, and cultural information (Dempsey, 2005; Jordan, 2004). Outside forces may influence families to take up new media and associated lifestyles and result in changes in children's reading, studying, and leisure activity patterns. Any association between media patterns and reading, studying, and socializing of children may be spurious. Therefore, it is important to control for factors that may influence both media use and children's other activities, including socioeconomic status (SES), maternal education and reading ability, race/ethnicity, number of parents, maternal employment, and number of children. Unlike other behaviors, displacement effects cannot be examined over time because normal development-related activities change as children move into adolescence. At one point in time, some kinds of activities are associated and others are not for individual children, controlling for confounding variables. Negatively associated activities are incompatible; positively associated activities or those with no association are compatible.

Because media use is selective, it is also important to control for initial levels of achievement and behavior in examining outcomes over time. If aggressive children attend to more violent media than less aggressive children; media use may have no additional impact. Additionally, parents who fear negative effects may intervene in advance, thus limiting potential harm (Nikken & Jansz, 2006). A more rigorous test of causality is whether a *change* in media use over time is associated with *change* in behavior or achievement for that same child.

Hypotheses

Displacement of healthy activities. The first hypothesis explored here, from displacement theory, is that the more time spent on media, the less time spent in all other activities considered to be valuable for children's health and well-being, including reading for pleasure, studying, nonscreen playing, sleeping, and playing sports or spending time outdoors. If media displace functional equivalents, then new media should displace old (TV), computer studying should displace noncomputer studying, and video or computer game play should displace each other. If structured displace less structured activities, then computer or video game play would displace outdoor activities, reading, nonscreen play, and sleep. If pleasure is the issue, then computer or video game play would replace noncomputer studying.

Observational learning. The second hypothesis is that computer use will facilitate learning such that children who use a home computer for studying, sending e-mail or text messages, or exploring the Internet will demonstrate higher vocabulary, reading, and problem-solving test scores. Similarly, greater video or computer game play will be associated with improved test scores, particularly on problem solving. However, the benefits of computer game and video game play are likely to accrue at low or moderate levels rather than at higher levels. Greater TV use is expected to be associated with higher achievement if it promotes learning but lower achievement if it displaces studying and reading. Prediction is weak because this study is unable to distinguish educational from entertainment viewing.

Cultivation theory: Behavior problems. The third hypothesis is that greater computer game and video game play will be associated with more behavior problems such as social withdrawal. High video game play will also be associated with more aggressive behavior problems because of its violent content. Finally, greater computer use for communications such as e-mail, instant messaging, and Internet use will be associated with fewer behavior problems because of its positive social networking function.

Use and motivation theory. The fourth hypothesis is that the effects of computer and video game use on achievement and behavior will differ by the race/ethnicity, age, and gender of the child. For example, Black children's achievement is expected to benefit from computer use more than that of their White peers. Younger children may be more influenced by computers and games than older children. Finally, because of their greater selection of violent games and greater salience to their peer group, boys' behavior may be more strongly linked to their computer and video game use, whereas girls' achievement may be influenced more by variation in computer and video game use because of their lower average usage.

Method

Data

The study sample consisted of 6- to 12-year-old children living with their mothers. The data were drawn from the Panel Study of Income Dynamics (PSID), a 30-year longitudinal survey of a representative sample of U.S. men, women, children, and their families. In 1997 the PSID added a refresher sample of immigrants to the United States. Also in 1997, with funding from the National Institute of Child Health and Human Development, data were collected on up to two randomly selected 0- to 12-year-old children of PSID respondents both from the primary caregivers and from the children themselves, the Child Development Supplement (CDS). The first CDS survey period began in March 1997 and ended in early December 1997 with a break from mid-June through August; thus, the study took place only during the spring and fall. Interviews were completed with 3,563 children in 2,380 child households with a response rate of 88%. From fall 2002 through spring 2003, the participants in the 1997 CDS were recontacted and another supplement administered. The total potential number of children eligible in 2002–2003 was 3,191, of whom 2,907 (91%) participated in the second CDS. When weights are used, the PSID is representative of U.S. individuals and their families (Fitzgerald, Gottschalk, & Moffitt, 1998).

For the descriptive analysis of activities and for the displacement analysis, the sample used in the present study consisted of 1,951 boys and girls between 6 and 12 years of age in 1997, of whom 1,871 lived with their mothers, and 1,607 boys and girls who were aged 6–12 in 2003, of whom 1,581 were living with their mothers. The 1997 and 2003 CDSs collected complete time diaries for one weekday and one weekend day for 1,448 (77%) of the 1,871 children aged 6–12 in 1997 and 1,317 (83%) of the 1,581 children aged 6–12 in 2003 living with their mothers. Comparisons between children who provided a diary and those who did not showed no significant differences on demographic characteristics.

For the outcome analysis, a longitudinal sample of children 6–12 years of age in 1997 who were reinterviewed in 2003 when they were between 12 and 18 years of age was created. Of the 1,448 children interviewed in 1997 and eligible, 1,226 were reinterviewed in 2003.

Measures

Time Use

The time diary, which was interviewer administered either to the parent or to the parent and child, asked about the child's flow of activities over a 24-hr period beginning at midnight of the randomly designated day. The questions asked the primary activity that was going on at that time, when it began and ended, and whether any other activity was taking place. The full instrument can be seen at http://psidonline.isr.umich.edu/CDS/TDqnaires.html. Activities need to occur regularly to be reliably reported in a diary. The activities selected occur daily (sleeping, nonscreen play, watching TV) or several times per week (studying, reading, video game playing, computer use, outdoor activities, and sports).

Professionally trained staff employed by the data collection organization coded the data to a level of reliability exceeding 90%. Because of extensive follow-up and editing, diaries were complete. Time spent traveling for the purpose of engaging in a specific activity was included in that category but 2 children who were traveling throughout their 2 diary days were excluded. Secondary activities were not included in the analysis presented here. This ensured that time added to 24 hr, making estimates of children's activities exhaustive, mutually exclusive, and additive. Almost no secondary computer or video game use was reported. The same analyses were run using total TV use, primary plus secondary. The results and implications were the same as reported for primary time.

For comparison across time, the primary activities of children were classified into the 18 major categories used by Timmer et al. (1985) in the early 1980s and by Hofferth and Sandberg (2001a, 2001b). Computer time was separated from its original categories such as play, study, and hobbies. Because it could be educational or recreational, computer use was further divided into three categories: playing games, studying, and computer communications. The latter includes activities such as Internet searching, accessing Web sites, e-mailing, and instant messaging. Electronic video game time (such as on Game Boy and other hand-held video game devices) was also separated from the larger *play* category; the remaining noncomputer, nonvideo game play is referred to here as *nonscreen play*. This analysis focuses upon 10 common activities among children: watching TV, nonscreen play, playing video games, playing on the computer, studying using the computer, computer communications, noncomputer studying, reading for pleasure, playing sports, informal outdoor activities, and sleeping. Time spent attending school was used to adjust for the potential amount of nonschool leisure time available to a child. Specific codes used to define each of these categories in

1997 and 2003 are available from the author. The total time children spend in an activity is a function of the proportion who engage in the activity and the time participants spend in it. An estimate of weekly time was computed by summing weekday time (including those who do not participate and have zero time) multiplied by 5 and weekend day time multiplied by 2.

As an alternative measure of time in a media activity, time spent on the computer and playing video games was divided into four categories: no use, low use (under the 25th percentile), moderate use (25th to under the 75th percentile among users), and high use (75th percentile and up). To illustrate, the 25th percentile for computer users was 1.6 hr and for video game users it was 2 hr. The 75th percentile for computer users was 6 hr and for video game users it was 7.2 hr. The results for low and moderate use were similar so they were combined. Because everyone watched TV, viewing hours were divided into low (lowest quartile: under 6.7 hr), moderate (25th–75th percentile: 6.75 to under 18.8 hr per week), and high (75th percentile and up: 18.8 hr plus).

Children's Behavior and Achievement

Children's *socioemotional adjustment* was measured by the Behavior Problems Index, a parent-reported measure of the incidence and severity of child behavior problems using 30 items originally drawn from the Achenbach scale (Peterson & Zill, 1986). This same scale has been used in the Child–Mother study of the 1979 National Longitudinal Survey of Youth since 1986 (Rogers, Parcel, & Menaghan, 1991). Using factor analysis with oblique rotation to permit interfactor correlation, two sets of factor scores, one with 16 items measuring withdrawn or distressed behavior (fearful, anxious, sad, withdrawn), called "internalizing," and one with 14 items measuring aggressive and impulsive behavior (disobedient, argumentative, moody, destructive), called "externalizing," were created. The correlation between these two factors was .51. For our longitudinal sample of 1,226 children whose families were interviewed in both 1997 and in 2003, the alphas were .88 for externalizing and .86 for internalizing each year. The alpha for the total behavior problems scale was .92. After deleting 152 cases missing data on controls, 1,074 boys and girls were included in the behavior problems analysis.

Three subtests of the Woodcock–Johnson Revised Test of Basic Achievement measured *cognitive achievement*: letter–word identification, a test of children's ability to identify and respond to letters and words; passage comprehension, a test that measures reading comprehension skills; and applied problems, a test of skill in analyzing and solving practical numerical problems (Woodcock & Mather, 1989). These were administered in the child's home in 1997 and 2003 by trained interviewers using materials prepared and distributed by the publisher of the test. Mothers were also administered the passage comprehension test. Children had to have a test score in both years to be included. Scores were calculated based on national norms that were age-standardized with a mean of 100 and a standard deviation of 15. Not all eligible children were assessed; therefore, sample sizes were smaller for achievement than for behavior problems. A significantly higher proportion of the children who were not assessed were White because most of such families were in the Midwest, far from population centers where interviewing teams were located. Children who were not assessed were also slightly less likely to live with two parents. Race and number of parents were controlled in the analysis. There were no significant differences on other variables, including media use and other activities. The results suggest that the exclusion of these children from the analysis of achievement did not bias the representativeness of the sample. The sample size was 880 for passage comprehension, 901 for the letter–word test, and 898 for applied problems.

Control Variables

Besides the overall analyses by gender, age, and race/ethnicity, this study examined media use controlling for key demographic characteristics of the family: maternal education (no college, some college, college degree), total family income or the ratio of family income to the poverty line, maternal employment (employed vs. not employed), family structure (1 vs. 2 parents), and family size (1 or 2 vs. 3 or more children). Controls included season of interview (fall = September through November, winter = December through February, spring = March through June) and warm versus not warm state. All the definitions were consistent across the two waves of data except that of maternal employment. In 1997, maternal employment was defined as ever employed in the previous year, whereas in 2003, maternal employment was defined as employed at the time of the survey. The core PSID wave that collected employment information was conducted in 2001 and not in 2002; employment at the survey date was deemed to be a better indicator than employment more than a year prior to the survey. Although included in the displacement analysis, because it had substantial missing data and was never related to any of the child outcomes in the preliminary analyses, maternal employment was dropped in the longitudinal analyses of child achievement and behavior. The difference analysis included a set of dummy variables for whether a change in family structure had occurred (one parent in both years, one parent in 1997 and two in 2003, two parents in 1997 and one in 2003, and two parents both years—the comparison group).

Analysis Plan

Displacement Analysis Using Pooled Sample of Children 6–12

For the displacement analysis, the sample of children 6–12 in 1997 was pooled with the sample 6–12 in 2003, a total of 2,765 children. After deleting 203 cases missing on either date of interview or on maternal employment, there were 1,323 children 6–12 in 1997 and 1,239 children 6–12 in 2003, a total of 2,562 children in the pooled multivariate analysis. All analyses were weighted using population weights provided by the PSID–CDS adjusting for the original probability of selection and for attrition, which were then normalized on the sample. Robust standard errors were computed using Stata to adjust for clustering of children within families and across years. A dummy variable was included to control for the data collection year.

Instead of dividing time into categories (Attewell et al., 2003) or using ordinary least squares regression (OLS; Vandewater, Bickham, & Lee, 2006), the analyses of amount of time spent in the activity were based upon Tobit regression models. These models adjust for the fact that not all children engage in each activity (Tobin, 1958), but they permit keeping time at the interval level. If OLS were used, the regression slope would be biased by the inclusion of zero values. The Tobit coefficients reflect both the effect of the independent variable on the probability of the activity and on the hours spent in the activity by participants (McDonald & Moffitt, 1980). The higher the proportion of children who participate in the activity, the more the results reflect the hours among participants and thus the more similar the results become to those from OLS regressions just on participants. Therefore, for activities in which all or almost all children participate (sleep, play, TV), OLS was used.

Child Achievement and Behavior in the Longitudinal Sample

We conducted two types of longitudinal analyses on a subsample of children who were 6–12 in 1997 and who were interviewed 5–6 years later. Because initial analyses showed significant interactions between media use and gender, all analyses are shown by gender. The first analysis is a lagged dependent variable analysis in which the achievement or behavior outcome of boys (or girls) in 2003 was regressed on media use in 1997, controlling for the same achievement or behavior measure in 1997 and a set of other control variables using OLS. These analyses control, in addition, for the mother's passage comprehension score because previous research has shown that children's scores are linked to maternal reading and vocabulary ability (Shonkoff & Phillips, 2000). The results can be interpreted as the change between 1997 and 2003 in the test score of a child who reported 1 more weekly hour of computer use, video game play, or TV viewing in 1997 compared with a comparable child who reported 1 fewer hour. Interactions between Black race/ethnicity and computer play, computer communications, or video game time are included to test whether such media use differentially affects children of minority racial/ethnic groups. There were too few cases to interact Black with computer study or Hispanic with any of the media variables. Because the ratio of family income to needs was not related to behavior and was related only to the letter–word score of boys (in the unexpected direction), interactions between income and computer or video game use were not examined. An examination of the interaction between age group (6–8, 9–12) and media use showed a consistent interaction between age group and computer study because few 6- to 8-year-olds used the computer for studying. This was not a substantively important finding, and age interactions were not pursued further. The second model examines whether the effects of computer, video game, and TV time are nonlinear, differing for moderate and high users, compared with low users or nonusers.

Our second analysis is a difference analysis. In this more rigorous test, the outcome is the change in achievement or behavior of the individual child across time associated with a change in media time.

Differences in achievement and behavior problems scores were regressed on the difference in hours spent on the computer, playing video games, and watching TV between 1997 and 2003 (2003 minus 1997). Because variables that do not change over time are not included, the only control variables included are age, number of children, household income, and family structure (in dummy categories). The results can be interpreted as the change in child achievement or behavior test score associated with spending an additional hour of time on the computer, video game, or TV between 1997 and 2003, net of the controls. Results are presented separately for White boys, Black boys, White girls, and Black girls. Sample sizes were too small for separate analyses of Hispanic children. Effect sizes (*d*) were calculated by dividing the coefficient by the standard deviation of the dependent variable (Cohen, 1988).

Results

Univariate and Bivariate Findings

Sample Characteristics

To examine time spent in activities in 1997 and 2003, the pooled cross-sectional sample of 2,562 children was used. Of the children in this sample, 41% were 6–8 and 59% were 9–12 (not shown). The sample was evenly split by gender. Three fourths lived with two parents, 44% of the mothers had completed at least some college, and two thirds were employed. Seventy-two percent were White, with 15% Black and 13% Hispanic. The average ratio of income to poverty was 3.4. Most children lived in nonwarm states; only one third of children lived in a warm state. Twenty-seven percent of the interviews were conducted in the fall, 41% in the spring, and 32% in the winter.

Bivariate Analyses of Children's Activities in 1997 and 2003

Columns 1 and 4 of the upper panel of Table 1 show the percentage of all children 6–12 who engaged in each of the media activities (watching TV, using the computer, and playing video games) and other activities (nonscreen playing, sleeping, reading, nonscreen studying, playing sports, engaging in outdoor activities, and spending time in school) in the interview week in 1997 and 2003. Columns 1 and 4 of the lower panel show the average weekly hours spent in each activity each year, including those who spent no time in the activity. Comparable estimates for boys and girls by year are shown in columns 2, 3 and 5, 6. *T* tests were conducted to examine change across time among all children (column 7), change across time by gender (columns 8 and 9), and gender differences within each year (columns 10 and 11).

Overall media activities. In 1997 and 2003 almost all children watched TV (94.8% to 96.6%; Table 1, upper panel). Between 1997 and 2003 the viewing time of 6- to 12-year-olds increased from 13.09 to almost 14 (13.96) hr per week (Table 1, lower panel). The proportion of children 6–12 years old who used the computer in the study week increased from 18.2% to 29.2%. In 2003, children 6–12 averaged about 1.33 weekly hours on the computer, including those who did not spend any time on it.

Screen game play. Computers were primarily used for playing games, a use that increased over the period. The proportion of 6- to 12-year-olds playing games on the computer rose from 14.1% in 1997 to 21.0% in 2003 and the average time (including zeroes) spent playing games increased from 0.58 to 0.81 hr per week. In 2003, younger children were as likely as older children to play games on a computer, 21.9% of 6- to 8-year-olds and 20.3% of 9- to 12-year-olds (not shown). The proportion of children 6–12 playing video games also increased between 1997 and 2003. In 2003, 36.1% of 6- to 12-year-olds played video games, compared with 29.6% in 1997. In 2003, they spent about 2.29 hr a week doing so, compared with 1.52 hr in 1997. The proportion playing video games in 2003 was slightly higher for 9- to 12-year-olds than for 6- to 8-year-old children (39.3% vs. 31.7%, not shown). Gender differences in video game play, in contrast, were large. In 1997, boys were more than twice as likely to play video games (40.7%) as girls (18.5%). In 2003, the proportion of boys playing video games was 56.9% compared with 16.5% of girls and the only significant gender difference in media time was the time spent playing video games. Girls spent 0.82 hr and boys spent 3.86 hr playing video games in 2003.

Nonscreen play. Three fourths of children (73.9%) engaged in nonscreen play. In 2003, the total weekly time 6- to 12-year-olds spent in nonscreen play was 6.33 hr (Table 1, lower panel), representing a 20% decline from 7.85 hr in 1997. Because time spent in computer and video game play increased, total play time (sum of nonscreen, computer, and video game play) declined only slightly,

Table 1

Percentage of Children 6–12 Participating in Selected Activities and Weekly Time, 1997 and 2003, by Gender

	1997			2003			1997 versus 2003			1997	2003
Activities	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	Boys versus girls	Boys versus girls
N	1,323	671	652	1,239	637	602					
Percent participating											
TV, total	94.8%	95.5%	94.2%	96.6%	97.5%	95.8%	*				
Computer, total	18.2%	18.0%	18.4%	29.2%	28.6%	29.8%	***				
Computer play	14.1%	14.6%	13.7%	21.0%	22.4%	19.6%	***				
Computer study	2.0%	1.5%	2.5%	2.6%	2.2%	3.0%					
Computer communications	4.7%	4.9%	4.5%	10.8%	8.7%	12.8%	***		*		*
Video games	29.6%	40.7%	18.5%	36.1%	56.9%	16.5%	**	***	***	***	***
Nonscreen play	83.4%	81.9%	84.9%	73.9%	76.0%	71.8%	***				
Sleeping	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%					
Reading	38.3%	35.3%	41.5%	47.2%	45.1%	49.1%	***	*		*	
Nonscreen study	58.2%	61.5%	54.9%	65.6%	65.9%	65.3%	***	*		*	
Sports	76.2%	79.2%	73.1%	60.4%	63.4%	57.5%	***	**	*	**	*
Outdoors	16.2%	16.9%	15.6%	10.0%	7.9%	12.0%	***		*		*
School	90.8%	91.6%	89.9%	90.6%	90.9%	90.2%	***				
Weekly hours of everyone (users and nonusers) in activities											
TV, total	13.09	13.19	12.99	13.96	14.14	13.80	*				
Computer, total	0.82	1.02	0.61	1.33	1.37	1.29	***	**		**	
Computer play	0.58	0.73	0.42	0.81	0.92	0.72	**	**		**	
Computer study	0.10	0.10	0.10	0.09	0.06	0.11					
Computer communications	0.14	0.20	0.09	0.43	0.39	0.46	***	*		*	
Video games	1.52	2.29	0.75	2.29	3.86	0.82	***	***	***	***	***
Nonscreen play	7.85	7.47	8.24	6.33	6.38	6.27	***				
Sleeping	69.03	68.86	69.20	70.74	70.51	70.96	***				
Reading	1.19	1.09	1.28	1.56	1.43	1.68	***				
Nonscreen study	2.91	3.15	2.66	3.57	3.57	3.56	***	*		*	
Sports	5.84	7.07	4.60	3.95	4.67	3.28	***	***	***	***	***
Outdoors	0.63	0.55	0.72	0.44	0.28	0.59	***		**		**
School	32.89	33.39	32.38	33.07	33.08	33.07					

Note. All data are weighted.
*$p < .05$. **$p < .01$. ***$p < .001$. t-tests across years for all and by gender, then by gender within year.

from 9.95 to 9.43 hr (not shown in table). The fraction of total play time that was represented by computer and video game play increased over the period relative to nonscreen time, from 21% to 33%.

Communications. Although still small, the use of the computer for communications, particularly the Internet and e-mail, more than doubled between 1997 and 2003, from 4.7% to 10.8%. The time spent in such activities rose from 0.14 hr in 1997 to 0.43 hr. In 2003, girls were more likely than were boys to use the computer for communications activities, such as e-mail and surfing the Internet: 12.8% versus 8.7%, respectively.

Studying and reading. The proportion of children who reported spending time studying increased significantly between 1997 and 2003. The increased studying documented here did not result from increased use of the computer for homework. Sixty-six percent of 6- to 12-year-old children reported nonscreen study time in 2003, compared with 58.2% in 1997, an increase of 14%. The total weekly time spent in nonscreen study increased from 2.91 hr to 3.57 hr for all 6- to 12-year-olds. The proportion studying with the computer (2% to 3% of children aged 6–12) and the amount of time spent studying on the computer (0.09–0.10 hr) did not change over the period and both were low.

Increases occurred in reading over the period. In 2003, 47.2% of 6- to 12-year-olds reported reading for fun during the survey week compared with 38.3% in 1997, an increase of 24% over the period. There was also a 31% increase in reading time for the entire age group—from 1.19 to 1.56. Thus, there has been a real increase in reading, though, overall, the proportion and the amount of time spent reading for pleasure were low.

Physical activities and sleep. Sports include formal team sports and informal sports activities for recreation. Outdoor activities include activities such as walking, pleasure drives, horseback riding, gardening, and camping. Participation in sports was lower in 2003 (60.4%) than in 1997 (76.2%). The amount of time spent playing sports was also lower by almost 2 hr in 2003 (3.95) compared with 1997 (5.84). Participation in outdoors activities declined by more than 50% from 1997 (16.2%) to 2003 (10%), and the amount of time spent in such activities also declined over the period, from 0.63 to 0.44 hr. All children sleep, but they may sleep for different amounts of time. The results indicate that time spent sleeping increased slightly from 1997 to 2003 for 6- to 12-year-olds from 69.03 to 70.74 hr per week.

Trade-Off Between Media and Other Activities

The descriptive results show gross change in activities over time; however, they do not adjust for potential differences in family socioeconomic circumstances and seasonality across the 2 years. The bottom panel of Table 2 shows the trade-offs between children's video game time, TV viewing time, and computer time (in the rows) and the key dependent variables of time in nonscreen play, sleep, reading, nonscreen study, sports, and outdoor activities in columns across the top, adjusting for SES and other differences (shown in top panel). The table also includes TV viewing, video game play, computer play, computer study, and computer communications across the top to test cross-media trade-offs. Hours spent in school are controlled to adjust for the hours they have available. Results for boys and girls were similar and are not presented separately.

Is Greater Media Use Associated with Less Time in Unstructured Activities Such as Nonscreen Play and Sleep?

More time on the computer (playing, studying, or communicating), playing video games, or watching TV were all associated with reduced time spent in nonscreen play. The results also showed that greater amounts of time spent playing computer games, using the computer for communicating, and watching TV were associated with less sleep. Neither computer studying nor video game playing was associated with reduced sleep.

Is Reading for Pleasure Displaced by Media?

The more time spent playing video games and watching TV, the less time spent reading for pleasure. However, spending more time in computer play, studying, or communicating was not associated with reduced time reading. In fact, the more hours spent using the computer for studying, the *greater* the time spent reading.

Is Studying Being Displaced by Media?

As with reading, the results showed that greater video game and TV time were associated with less nonscreen study time. The amounts of time spent studying with or communicating using the computer were not associated with time spent in nonscreen studying. However, more time spent playing computer games was associated with less nonscreen study time.

Does Media Time Interfere with Playing Sports or Spending Time Outdoors?

The associations between TV viewing hours and sports time and between computer play and sports time were negative. Video game time, computer study hours, and computer communications were not related to time playing sports. Greater video game time was associated with less time spent in *outdoor activities*, but computer time was not.

Are There Trade-Offs Among the Different Forms of Media?

Video game play and TV time do not appear to be incompatible. Children who spent more time playing video games did not appear to spend less time watching TV; however, there was a negative association between computer studying and TV viewing. Finally, video game time was negatively associated with computer game play and computer communication hours.

Media and Achievement, Lagged Analysis

Table 3 provides the means and standard deviations for the longitudinal file of children 6–12 in 1997 who were reinterviewed in 2003. Black children lived with mothers with less education and a lower vocabulary score, their family was more likely to be headed by a single mother, and their family incomes were lower than those of White children. Black children had lower achievement test

Table 2

Regression Coefficients for Trade-Off Between Weekly Hours in Selected Activities

	Nonscreen play	Sleeping	Reading	Nonscreen study	Sports	Outdoors	TV	Video games	Computer play	Computer study	Computer communications
Variable	OLS	OLS	Tobit	Tobit	Tobit	Tobit	OLS	Tobit	Tobit	Tobit	Tobit
Background											
Black	−2.06**	0.53	−1.29**	1.37**	−2.22**	−3.91**	2.01**	0.66	−0.98	−3.53	−5.41***
Hispanic	−1.78**	−0.05	−1.50*	1.89**	−2.29*	−1.11	1.02	−1.94	−2.50*	−1.57	−0.65
Male	0.32	0.00	−0.11	0.72	2.75***	0.09	0.39	7.24***	1.25*	−2.14	−0.02
Age, 6–8	3.31***	2.56***	0.00	−2.44***	−2.33***	0.14	−1.79**	−1.83**	−1.28*	−4.82**	−4.23***
Mother completed some college	−0.10	−1.17*	−0.02	0.37	0.14	0.89	−3.16***	−0.27	1.38	−0.49	1.93
Mother completed college or more	−0.86	−1.18*	0.99*	0.82	1.28*	0.94	−3.47***	−2.29**	1.26	2.21	0.72
Mother is employed	−1.66***	−0.24	−0.62*	0.21	0.81	−0.51	−0.26	0.26	−0.48	−0.71	0.44
Two parents (vs. one parent)	0.24	0.66	0.48	−0.40	−1.47*	0.83	−0.49	−0.48	0.35	−0.75	−0.40
Three or more children	0.41	0.52	0.14	−0.46	−0.01	−0.55	−0.41	−0.58	0.21	−1.11	−0.09
Family income to poverty ratio	0.03	−0.11	0.05	0.05	0.05	−0.10	−0.18**	0.04	0.09*	0.04	0.09
Lives in warm state	0.30	−0.61	−0.34	0.15	1.42*	0.40	−0.26	−1.33*	−1.58*	1.83	−1.96*
Interview conducted in fall	0.40	−0.68	−0.10	−0.20	0.05	0.14	−1.62*	−1.30	−0.71	1.14	−0.40
Interview conducted in spring	0.86	−1.88*	−0.24	−1.86**	2.54*	1.59	−0.68	0.18	−0.49	−1.13	−0.68
Year, 1997	1.06	−0.85	−0.85	−0.18	0.95	1.36	−0.45	−2.35*	−1.61	0.37	−3.64
Weekly hours											
Computer play	−0.19**	−0.18*	−0.04	−0.22*	−0.26*	−0.23	−0.18	−0.11	—	0.36	0.44***
Computer study	−0.49***	−0.41	0.26*	−0.05	−0.21	0.14	−0.56**	−0.14	0.64*	—	0.69*
Computer communications	−0.21**	−0.30**	−0.10	0.15	−0.20	−0.05	0.02	−0.46*	0.20	0.57*	—
Video games	−0.20***	−0.08	−0.16***	−0.12**	0.00	−0.42*	−0.06	—	−0.16*	−0.04	−0.28*
TV	−0.11***	−0.18***	−0.10***	−0.11***	−0.16***	−0.08	—	−0.03	−0.06	−0.28***	−0.01
School	−0.16***	−0.25***	−0.06***	0.08***	−0.10***	−0.05	−0.22***	−0.05	−0.07**	−0.03	−0.01
Constant	13.29***	81.74***	3.26***	1.69†	7.07***	−8.70***	24.95***	−1.82	−2.61†	−12.78**	−7.13**

Note. $N = 2{,}562$. Data are weighted. Robust standard errors computed using Stata. OLS = ordinary least squares.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 3

Means of All and Standard Deviations of Continuous Variables in Longitudinal Analyses

	Girls							Boys						
	Total		White		Black		*T* test White versus Black	Total		White		Black		*T* test White versus Black
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	*M*	*SD*	
Background														
White and other	0.73	—	—	—	—	—		0.72	—	—	—	—	—	
Black	0.13	—	—	—	—	—		0.18	—	—	—	—	—	
Hispanic	0.14	—	—	—	—	—		0.10	—	—	—	—	—	
Age of child	9.56	2.05	9.64	1.99	8.94	2.18	**	9.32	1.99	9.33	1.95	9.54	2.09	
Mother completed high school or less	0.53	—	0.41	—	0.71	—	***	0.49	—	0.41	—	0.59	—	**
Mother completed some college	0.25	—	0.31	—	0.19	—	*	0.28	—	0.30	—	0.32	—	
Mother completed college or more	0.22	—	0.28	—	0.10	—	***	0.23	—	0.29	—	0.08	—	***
Mother's passage comprehension score	31.73	5.17	33.19	4.46	27.39	5.34	***	32.50	4.79	33.87	3.73	28.49	4.50	***
Two parents (vs. one parent)	0.81	—	0.87	—	0.37	—	***	0.77	—	0.85	—	0.38	—	***
Three or more children	0.46	—	0.39	—	0.51	—		0.45	—	0.39	—	0.50	—	*
Family income to poverty ratio	3.31	4.46	4.02	4.99	1.42	1.49	***	3.28	3.07	3.81	3.15	1.80	1.76	***
Dummy variables for media use														
Computer hours (low to moderate)	0.17	—	0.20	0.40	0.06	0.24	**	0.12	0.32	0.15	0.36	0.04	0.20	**
Computer hours (high)	0.03	—	0.03	0.18	0.01	0.12		0.07	0.25	0.08	0.27	0.04	0.19	
Video game hours (low to moderate)	0.15	—	0.16	0.37	0.11	0.31		0.32	0.47	0.33	0.47	0.31	0.46	
Video game hours (high)	0.02	—	0.02	0.13	0.07	0.26	***	0.10	0.30	0.10	0.30	0.13	0.34	
TV hours (moderate)	0.02	—	0.03	0.17	0.01	0.08		0.04	0.20	0.05	0.22	0.02	0.15	
TV hours (high)	0.24	—	0.19	0.40	0.40	0.49	***	0.22	0.42	0.20	0.40	0.37	0.48	***
Achievement and behavior														
Passage comprehension 1997	108.59	16.53	110.53	15.98	99.75	16.09	***	105.71	15.88	108.09	15.41	96.83	13.19	***
Passage comprehension 2003	103.75	15.11	107.15	14.22	94.95	12.81	***	101.65	17.31	106.09	16.84	91.40	14.06	***
Applied problems 1997	109.13	16.60	111.59	15.71	98.44	17.48	***	112.86	18.47	116.50	17.72	100.22	12.78	***
Applied problems 2003	102.58	17.12	106.92	15.70	90.75	13.62	***	105.83	16.56	110.55	15.85	93.44	11.74	***
Letter–word 1997	109.87	19.14	112.18	18.72	99.31	15.09	***	105.96	19.27	108.64	18.47	95.68	18.41	***
Letter–word 2003	107.47	20.45	111.11	20.21	93.68	16.66	***	103.11	20.37	108.37	19.87	89.10	16.24	***
Total behavior problems 1997	39.64	8.45	39.93	8.81	38.04	8.07		40.77	8.22	40.37	7.92	41.48	8.83	
Total behavior problems 2003	41.04	9.26	40.91	9.3	39.55	8.1		40.91	9.31	40.54	8.89	41.37	10.62	
External behavior problems 1997	0.13	1.02	0.17	1.07	−0.19	0.80		0.10	1.03	0.17	1.09	−0.01	0.95	
External behavior problems 2003	0.08	1.00	0.04	0.99	−0.19	0.79		0.01	1.00	−0.02	0.76	0.07	1.08	
Internalizing behavior problems 1997	−0.24	0.88	−0.23	0.90	−0.27	0.86		0.02	1.02	−0.04	1.02	0.26	1.04	
Internalizing behavior problems 2003	−0.12	0.92	−0.14	0.92	−0.09	0.85		0.08	1.06	0.04	1.04	0.28	1.22	
N	538		297		201			536		288		218		

Note. All data are weighted. Hispanic children are included in totals.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

scores than did White children but a similar level of behavior problems. Finally, Black boys and girls were less likely to use the computer at home during the study week but were more likely to watch a lot of TV than were White boys and girls.

Effects of Background Factors on Achievement in 2003

Table 2 showed that, compared with those whose mother had not attended any college, children whose mothers had completed college played fewer video games, watched less TV, and spent more time reading. Similarly, those living in families with a higher ratio of income to needs watched less TV and played more on the computer. These SES–activity links would likely produce a positive association between computer use and achievement and a negative association between video game use and achievement if SES were not controlled.

Main Effects of Study Time on Achievement in 2003

Initial analyses showed significant zero-order associations between computer and video game time and children's test scores, However, when background factors and the child's and mother's 1997 score on that same test were controlled, there were no associations between computer game play hours, computer communications, video game hours, or TV viewing hours in 1997 and achievement in 2003 for boys or girls (Table 4). Two associations were significant. For girls (Table 4, Model 1), greater computer study hours were associated with a significantly lower score on the applied problems test, opposite that expected ($b = -2.31, p < .05$). The effect size (d) was −.14 ($p < .05$) for girls' and −.02 (*ns*) for boys' computer study. Pooling over boys and girls (not shown), the interaction between gender and time spent studying on the computer was significant. Girls who spent more time studying had a significantly lower applied problems test score than boys who spent more time studying. There was also a significant negative association between studying and boys' scores on the passage comprehension test ($b = -0.71, p < .05$). However, this did not differ by gender. The coefficient for girls was about the same size as for boys but was not significant.

Main Effects and Interactions Between Race and Time Spent in Computer Communication

Although the main effects of computer communications and play were not significant, there were substantial main effects of race/ethnicity, with minorities achieving lower scores on all the achievement measures. Black boys, scored 7.77 points lower than White boys, and Black girls scored 3.57 points lower than those of White girls on the passage comprehension test. Interactions between Black race and computer use and Black race and video game use were included in Model 1 of Table 4. The main effects of computer study hours did not change and computer communication and computer play hours were still not significantly associated with the child's passage comprehension score, but the interaction between Black and computer communication hours was significant for several achievement tests. Black boys and girls who spent more time in 1997 in communications activities, such as surfing the Net and e-mail, had passage comprehension scores in 2003 higher by 8.98 and 8.33 points, respectively, than Whites who were similarly active. After adding together the main and interaction effects of Black race and communication to obtain total effects, Black boys' passage comprehension scores were 1.03 points above those of White boys ($d = .06$), and Black girls' passage comprehension scores exceeded those of comparable White girls by 5.92 ($d = .40$), if they used the computer for communication. Black girls also had higher scores than White girls on the letter–word test if they were more active users of computers for communication; the sum of the main and interaction effects was 1.12 ($d = .05$).

Amount of Use

Discrete categorical effects may not be captured by these continuous measures. Less than one fifth of elementary-school-age children used the computer at home during the week in 1997 but the top quartile spent 6 or more hr on the computer. The analysis of media use grouped by low to moderate and high versus no hours (or medium and high vs. low TV hours) is shown in Model 2 of Table 4. *Low to moderate computer hours* were positively and significantly associated with higher passage comprehension scores ($b = 4.12, p < .05, d = .27$) and with marginally higher applied problems scores ($b = 3.21, p < .10, d = .18$), but only for girls. *High video game hours* were, for girls, associated with a significantly lower score ($b = -11.04, p < .05, d = .73$) on the letter–word test and a marginally lower score ($b = -7.75. p < .10, d = .48$) on the applied problems test. There were no associations between the measures of boys' overall computer and video game time in 1997 and their achievements in 2003.

Table 4

Regression Coefficients for Child Scores in 2003 by Weekly Media Use in 1997, by Gender

Variable	Letter–word Model 1	Letter–word Model 2	Passage comprehension Model 1	Passage comprehension Model 2	Applied problems Model 1	Applied problems Model 2	Behavior problems Model 1	Behavior problems Model 2
Boys								
Computer play	0.36		0.50		0.33		0.02	
Computer study	−0.45		−0.71**		0.39		0.08	
Computer communications	−0.65		−0.18		−0.43		−0.42	
Video games	−0.05		0.02		0.01		−0.02	
TV hours	−0.15		0.08		0.07		0.01	
Child is black	−10.55**	−9.34**	−7.77*	−6.32*	−6.11**	−5.36**	−0.79	−0.73
Child Is Black × Computer Play	−1.02		−0.23		−0.77		0.67	
Child Is Black × Computer Communications	2.37		8.98*		3.50		1.63	
Child Is Black × Video Game	0.21		0.42		0.17		−0.18	
Computer hours (low to moderate)		0.68		2.13		3.65		−1.25
Computer hours (high)		−1.50		7.93		−2.19		−2.63
Video game hours (low to moderate)		0.36		0.21		−1.21		0.60
Video game hours (high)		0.84		1.44		2.07		0.13
TV hours (moderate)		2.79		−4.33		7.40		5.28
TV hours (high)		−2.23		1.67		1.23		0.98
R^2	.58	.57	.46	.45	.53	.53	.33	.32
N	439	439	428	428	439	439	536	536
Girls								
Computer play	1.21		0.31		0.31		−0.50†	
Computer study	−1.23		−0.68		−2.31**		−0.26	
Computer communications	−1.20		0.66		0.23		0.01	
Video games	−0.52		0.14		−0.11		0.33	
TV hours	0.01		0.08		−0.02		0.05	
Child is black	−3.67	−3.36	−3.57*	−3.20*	−5.81**	−5.61**	−0.72	−0.94
Child Is Black × Computer Play	−1.24		−0.82		−0.29		−0.02	
Child Is Black × Computer Communications	5.99*		8.33**		−3.23		0.29	
Child is black × video game	−0.56		0.41		−0.66		−0.12	
Computer hours (low to moderate)		3.06		4.12*		3.21†		−1.48
Computer hours (high)		12.02		16.50		6.20		−8.86***
Video game hours (low to moderate)		4.28		0.22		0.45		1.06
Video game hours (high)		−11.04*		0.14		−7.75†		1.93
TV hours (moderate)		−7.26		−20.83		−5.60		7.98**
TV hours (high)		0.58		0.52		−1.26		1.35
R^2	.54	.55	.40	.41	.50	.50	.35	.34
N	462	462	452	452	459	459	538	538

Note. Data are weighted. Robust standard errors computed using Stata. Controls include race/ethnicity, child age in 1997, maternal education, maternal passage comprehension score, number of parents, number of children, ratio of family income to poverty, and the child's test score in 1997.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Media and Behavior Problems, Lagged Analysis

Main Effects of Media and Media–Gender Interactions on Behavior Problems

We report results only for total behavior problems; similar results were obtained for internal and external problems. No direct associations between media use and behavior problems were found for boys (Table 4, Model 1). For girls, greater computer play hours were associated with marginally reduced behavior problems ($b = -0.50$, $p < .10$, $d = .05$). There was a significant interaction between gender and computer play. Girls who spent more time playing computer games had a lower behavior problems score than boys (not shown).

Interaction Between Media Time and Race

Media time and race did not interact in their association with behavior problems.

Amount of Media Use

Finally, in Model 2 of Table 4, which categorized media by level of use, spending a greater amount of time on the computer was associated with lower behavior problems for girls ($b = -8.86$, $p < .001$) but not boys, and the effect size was large ($d = .95$). Video game hours were not significantly linked to behavior problems for girls and none of these categories was linked to boys' behavior problems. Moderate TV viewing was associated with greater behavior problems for girls ($b = 7.98$, $p < .05$) but not boys.

Changes in Children's Test Scores, 1993–2003

Table 5 shows the results of regressing the difference between children's 1997 and 2003 achievement test and behavior problems scores on the difference

Table 5

Regression Coefficients for Changes in Child Test Scores Associated With Changes in Weekly Media Use Between 1997 and 2003

Variable	Letter–word	Passage comprehension	Applied problems	Internalizing problems[a]	Externalizing problems[a]
Non-Hispanic White boys[b]					
N	237	228	238	278	278
Change in weekly hours					
Computer play	0.09	−0.13	−0.24	−0.01	−0.01
Computer study	−0.23	−0.03	−0.33	0.00	−0.03
Computer communications	0.15	−0.27*	−0.28*	−0.01	−0.02
Video games	0.00	−0.03	−0.07	0.00	0.01*
TV	0.13*	0.01	0.03	0.00	0.00
Black boys[b]					
N	185	182	185	195	195
Change in weekly hours					
Computer play	0.98	0.10	−0.02	−0.03	−0.07
Computer study	0.63	3.71**	0.17	−0.36	−0.05
Computer communications	0.14	0.95*	0.53	0.01	−0.02
Video games	−0.09	−0.12	−0.08	0.00	0.01*
TV	0.10	−0.03	−0.05	0.01	−0.01
Non-Hispanic White girls[b]					
N	266	261	265	300	300
Change in weekly hours					
Computer play	0.55	0.99*	0.98**	−0.02*	−0.01
Computer study	−0.71	0.25	1.18*	0.05*	0.01
Computer communications	0.08	0.04	−0.02	−0.01	0.00
Video games	−0.69*	0.20	0.14	0.02	−0.02
TV	−0.01	−0.02	−0.06	−0.01	0.00
Black girls[b]					
N	179	175	177	188	188
Change in weekly hours					
Computer play	−0.19	−0.18	−0.28	0.05	0.02
Computer study	0.44	−0.47	−0.48	0.01	0.00
Computer communications	0.38	−0.22	−0.10	0.02	0.00
Video games	0.50	−1.22**	0.54*	−0.03	−0.01
TV	−0.04	−0.20**	−0.01	0.00	−0.01*

Note. Data are weighted. Robust standard errors computed using Stata.

[a] Standardized factor scores with mean of 0 and standard deviation of 1 used here.

[b] Controls for change in age of child, number of children, household income, and number of parents.

$^{*}p < .05$. $^{**}p < .01$.

between children's 1997 and 2003 media time. A positive association means that an increase in media time was associated with an increased score on that particular test. Because the previous analyses suggested differences both by gender and race, the table shows the results separately for White and Black boys and White and Black girls. The same background controls—age of child, number of children, household income, and number of parents—were included in each analysis, but are not shown. A table of effect sizes is available from the author.

Boys' Achievement

Among boys, the relation between computer use and achievement varied by race/ethnicity. For White boys, increased *computer communications* was associated with a significantly lowered score on both the passage comprehension ($b = -0.27$, $p < .05$) and applied problems ($b = -0.28$, $p < .05$) tests. For Black boys, in contrast, increased *computer communications* was associated with increased passage comprehension ($b = 0.95$, $p < .05$). For Black boys, in addition, increased *computer study time* was associated with an *increased* score on the passage comprehension test ($b = 3.71$, $p < .01$). Change in video game time was not associated with change in boys' achievement for any group. Finally, increased TV viewing hours were associated with an increased score on the letter–word test for White boys.

Girls' Achievement

Computer time was consistently positively associated with White girls' achievement. Greater time spent *playing games on the computer* was associated with increased passage comprehension ($b = 0.99$, $p < .05$) and applied problems test scores ($b = 0.98$, $p < .05$) between 1997 and 2003 for White girls.

In contrast to the positive association between computer game play and achievement, *playing video games* was detrimental to the achievement of White girls. For them, increased video game time was associated with a significantly reduced score on the letter–word test between 1997 and 2003 ($b = -0.69$, $p < .05$). For Black girls, increased video game hours were associated with a reduced score on the passage comprehension test ($b = -1.22$, $p < .05$) but with an increased score on the applied problems test ($b = 0.54$, $p < .05$). Increased TV viewing was also associated with a reduced passage comprehension score for Black girls.

Boys' Behavior Problems

For boys, there was no association between computer hours and behavior problems. In contrast, increased time spent playing video games between 1997 and 2003 was associated with increased externalizing behavior problems ($b = 0.01$, $p < .05$). Although the positive relation was found for both Black and White boys, the effect sizes were small, .01.

Girls' Behavior Problems

For White girls, playing more on the computer between 1997 and 2003 was associated with reduced internalizing problems whereas increased time spent studying was associated with increased internalizing behavior problems. Watching more TV was associated with fewer externalizing problems for Black girls. Finally, in contrast to boys, Black and White girls who played more video games between 1997 and 2003 did not experience increased internalizing or externalizing behavior problems.

Discussion

The research reported here described the use of media at home by U.S. 6- to 12-year-olds and linked it to their achievement and behavior in adolescence in 2003.

Change Over Time

Television remains the dominant medium in middle childhood. In 2003, children spent 14 hr per week watching TV, an increase from 13 hr in 1997. The use of computers and video games at home also increased between 1997 and 2003, but the proportion using them and the time spent were still relatively low. Almost 2 of 10 children 6–12 used the computer at home in an average week in 2003, and the average time spent was about 1 hr and 20 min over all children. Three fifths of children's computer time was spent playing games, one third was spent in communications, and the remainder was spent studying. Between 1997 and 2003 increases occurred both in computer game playing and in computer communications, primarily e-mail and Internet use. The small amount of time spent studying did not change. Although total computer time in 2003 was similar for boys and girls, girls were more likely to use the computer for communications and boys were more likely to use it to play games. Video game play was highly gender differentiated. In 2003, 57% of boys and only 17% of girls played a video game during the week of the interview. Including everyone, boys played for almost 4 hr compared with less than 1 hr for girls. This equals 6.8 hr for boys who played and 4.8 for girls who played. Because nonscreen play time declined for all children and computer and video game play increased, screen play occupied a larger fraction of children's total play time in 2003 than in 1997, 33% versus 21%.

Media Displacement of Other Activities

Some displacement was expected. As hypothesized, trade-offs across activities that are functionally equivalent, less structured, and less beneficial at the margin were supported. One example of the displacement of functionally similar activities was that greater video game play time was associated with less computer play time. There was no association between computer and noncomputer study time, indicating that computer studying is not equivalent to and does not displace nonscreen studying. Greater time in any media activity was associated with less time spent in nonscreen play, supporting the hypothesis of substitution of structured for less structured time. Greater use of all media except video games was associated with less sleep, supporting the marginal benefits hypothesis. Although TV time was also expected to decline, greater computer studying was the only activity associated with less TV viewing. The results are consistent with expansion in programming and inconsistent with a negative effect on old media of new media. TV has a unique niche—one that is not easily displaced.

Of new media, only video game play was associated with spending less time in three achievement- or health-related activities—reading, studying, and spending time out of doors. Computer play was also associated with less time spent in sports whereas video game play was not. The availability of sports games may make video gaming attractive to both sports participants and nonparticipants. Playing computer games was similar to playing video games in that each had a negative association with nonscreen studying.

Media Use and Achievement

Computer Use

Consistent with observational learning theory, greater *computer use* overall, particularly low to moderate computer use, appeared to be consistently and positively linked to reading and problem-solving achievement in adolescence 5 years later for girls but not boys. Use and motivation theory was important to the interpretation of the findings because the association depended on whether the computer was used for playing games, for studying, or for computer communications such as e-mail and Internet use. As suggested by use and motivation theory, associations varied by race/ethnicity as well as gender. For White girls, increased time spent playing games on the computer between 1997 and 2003 was associated with an increased score on the passage comprehension and applied problems tests in the change analyses. In the lagged regressions, Black girls who spent more time in computer communications had scores on the passage comprehension and letter–word tests high enough to outweigh their overall lower score compared with White girls. The overall positive effect of computer use for girls reflected the fact that the majority of computer time was spent playing games. This study, unfortunately, had no information on the types of games children were playing on the computer. However, girls of this age are likely to be playing educational games, in contrast to older children who may be playing games of strategy and adventure (Wright et al., 2001). Research indicates that girls are much less facile and comfortable with computers than boys (Vered, 1998); greater experience with them is likely to give girls an achievement edge.

For White boys the results were surprising. There were no positive associations of computer use with achievement in the lagged or difference models for White boys, and there was one negative association: An increase in time spent in computer communications between 1997 and 2003 was associated with a decline in passage comprehension and applied problems scores. Boys may become too involved in the e-mail and Internet functions of computers to focus on their homework assignments. Computers are not a magic bullet to improve children's academic success. Other research suggests that the use of computers for academic work is underdeveloped (Rathbun, West, & Hausken, 2003).

For Black boys, in contrast, the news was somewhat better. As for Black girls, in both the lagged variable and difference analyses, Black boys who spent more time in computer communications and studying between 1997 and 2003 had increased passage comprehension test scores. The results suggest that greater computer use is associated with higher achievement of White girls, Black girls, and Black boys.

Although a significant negative association between computer study and later achievement was found in the lagged dependent variable regressions for both boys and girls, this finding reversed direction in the difference analyses. The difference models found that increased computer studying was associated with an increased score on the applied problems test for White girls and on the passage comprehension test for Black boys. Study behavior is selective; students who are having problems in school may be encouraged to spend more time studying. The difference analysis examined changes in the achievement of the individual student associated with changes in media use across time, eliminating this selectivity.

Video Game Play

In contrast to what was expected based upon observational learning theory, video game playing had neither positive nor negative associations with any measures of boy's academic achievement. However, it had both positive and negative associations with girls' achievement. Even though fewer girls than boys played games, increased video game hours between 1997 and 2003 were associated with a lower letter–word test score for White girls and a lower passage comprehension score for Black girls. At the same time, increased video game hours were associated with an increased score on the applied problems test for Black girls. Video games do not generally improve practice in vocabulary or reading skills; however, they do teach problem-solving skills; the applied problems test is the one test on which Black girls improved their scores with increased video game play. The results support the argument that video game play has the potential to improve learning. However, excessive video game time appears detrimental to learning; a high level of video game play was linked to a reduced applied problems test score for girls in the lagged models. The reason for detrimental effects on learning is that video game time displaces valuable reading and study time. If reading and study time were not displaced, video games could improve learning. Lowered reading and studying time may be more detrimental to girls' reading achievement than that of boys. More research needs to consider the types and circumstances in which girls play video games and to examine differences between the game play of White and Black children.

Media Use and Behavior Problems

Cultivation theory focused upon the types of messages children receive from media. According to cultivation theory, greater computer and video game use should be associated with more externalizing behavior problems because of aggressive content and with more internalizing behavior problems because these are individual activities. Media may also be linked to behavior problems because of displacement of normal outdoor activities.

Computer Use

Consistent across both longitudinal analyses, increased computer game playing was associated with *reduced* internalizing behavior problems for girls, possibly because of social networking. However, consistent with the previous discussion about greater studying linked to school problems, the difference analysis found increased studying linked to *increased* internalizing problems for Black girls. Studying may have social costs in terms of isolation. Computer use did not influence boys' internalizing or externalizing behavior problems.

Video Game Play

Although computer time did not displace outdoor activities, video game play did. The hypothesis from cultivation theory predicting that greater video game use would be associated with behavior problems was supported for externalizing behavior for boys. In the difference analysis, increased video game hours between 1997 and 2003 were associated with increased externalizing behavior problems for all boys but not for girls. Boys who spend more time playing video games spend less time in active outdoor pursuits, which may be a source of these behavior problems. This finding supports other research that has found video game play to be associated with aggressive behavior; however, the effect sizes are quite small in this national sample of children. The results did not support the hypothesis that increased game play would lead to social isolation.

Variation in Media Influence by Race

Use and motivation theory led us to expect that there would be substantial variation in the association between media and children's achievement and behavior depending upon current use levels, motivation, and disposition of the user. Perhaps because of their lower initial level of computer use, the potential of home computer use for achievement was greater for Black boys and Black girls than White boys. Greater time spent with the computer for study and communications was associated with higher achievement for Black boys and Black girls compared with White boys. The boost resulting from computer communications alone was sufficient to offset the substantial Black–White test score gap. Black boys and White boys have a similar risk of increased aggressive behavior problems, but the risk is small. For African American children, the computer provides an opportunity for higher achievement with a minimal risk of internalizing or externalizing behavior problems.

Strengths and Limitations

There were several limitations of the present study. It examined only primary and not secondary media use. Cell phone and music media use, which are likely to fall in the multitasking category, have become widespread, requiring a method for examining how children spend their time that includes multiple uses of time. However, supplementary analyses indicated that incorporating multitasking would not have changed the findings. In addition, this study focused only on computer use at home and did not incorporate computer use in the school setting. Finally, although computer use was divided by type of use, TV viewing content was not examined. This limits the conclusions that can be drawn about the association of TV viewing with achievement. Consistent with this limitation, few associations between TV viewing time and children's achievement and behavior were found and the effects were small. The content of video games was also not considered; the potential association between violent video game time and behavior problems could not be examined.

This study has offsetting strengths. For the first time, this study provides a nationally representative picture of the primary media activities of children 6–12 in 1997 and 2003 in the United States, a description of the changes in media use that occurred in a period of rapid social change, and estimates of the association of changes in media use with changes in children's achievement from interviewer-administered standardized tests and behavior reported by parents.

Implications

This research shows that whereas TV and video games may crowd out positive activities, computer use does not. Home computer use among girls in middle childhood has potential benefits for their achievement in adolescence. The achievement of Black children is associated with greater computer use more than that of White children and there are relatively few risks to behavior for Black children. This research suggests that the vocabulary and reading test score gap between Black and White boys and girls may be at least partially offset through children's higher use of computers at home. Computer literacy is essential for full participation in the information society, and increased access to computers would reduce the disadvantage across generation. Although access and use at school could reduce this divide, most research suggests that, even today, only a fraction of children use computers regularly at school. At present, home computer use is limited among children from minority families; efforts to expand the presence of computers in low-income homes could improve their children's academic performance. The results showing greater potential benefits to Black children also suggest an area in which outreach could help narrow the achievement gap between Black and White children. Benefits also are likely to accrue to Hispanic children; the sample was too small for reliable estimates.

Finally, the conclusions regarding the positive associations of computer game play with achievement do not apply to video game play. Even though boys spend more time playing video games than girls, their achievement does not appear to be significantly affected, at least at the levels of play occurring in 1997 and 2003. Although girls' vocabulary and reading declined with greater video game use, increased video game play was associated with increased problem-solving skills for girls. As expected, moderate use appeared to be beneficial whereas extensive video game time was not. Boys' *behavior* is at risk as a result of video game use; the results showed a consistent detrimental association of video game time with the externalizing behavior of boys, though the effect was small. Greater video game time has been shown in other studies to lead to more aggressive behavior. This is an important caveat for those who write about the potential benefits of electronic game playing. More research is needed to include school use, to incorporate other media and multitasking, and to examine the longer term correlates of media involvement into young adulthood.

References

Alloway, N., & Gillbert, P. (1998). Video game culture: Playing with masculinity, violence and pleasure. In S. Howard (Ed.), *Wired up: Young people and the electronic media* (pp. 95–114). London: UCL Press.

Anderson, C., & Bushman, B. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior. *Psychological Science, 12*, 353–359.

Anderson, D., Fite, K., Petrovich, N., & Hirsch, J. (2006). Cortical activation while watching video montage: An fMRI study. *Media Psychology, 82*, 7–24.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, John. C. (2001). Early childhood television viewing and adolescent behavior. *Monographs of the Society for Research in Child Development, 66*(1).

Andreasen, M. S. (1994). Patterns of family life and television consumption from 1945 to the 1990s. In D. Zillmann, J. Bryant & A. Huston (Eds.), *Media, children, and the family* (pp. 19–36). Hillsdale, NJ: Erlbaum.

Attewell, P. (2001). The first and second digital divides. *Sociology of Education, 74*, 252–259.

Attewell, P., & Battle, J. (1999). Home computers and school performance. *The Information Society, 15*, 1–10.

Attewell, P., Suazo-Garcia, B., & Battle, J. (2003). Computers and young children: Social benefit or social problem. *Social Forces, 82*, 277–296.

Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 121–153). Mahwah, NJ: Erlbaum.

Chambers, J., & Ascione, F. (1987). The effects of prosocial and aggressive videogames on children's donating and helping. *Journal of Genetic Psychology, 148*, 499–505.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Coley, R., Cradler, J., & Engle, P. (1997). *Computers and classrooms: The status of technology in U.S. schools*. Princeton, NJ: Educational Testing Service.

Cordes, C., & Miller, E. (2000). *Fool's gold: A critical look at computers in childhood*. College Park, MD: Alliance for Childhood.

Dempsey, N. P. (2005). Television use and communication within families with adolescents. In B. Schneider & L. J. Waite (Eds.), *Being together, working apart* (pp. 277–296). Cambridge, England: Cambridge University Press.

Fisch, S. M. (2004). *Children's learning from educational television: Sesame Street and beyond*. Mahwah, NJ: Erlbaum.

Fisch, S., & Truglio, R. T. (2001). *"G" is for growing*. Mahwah, NJ: Erlbaum.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources, 33*, 251–299.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave/Macmillan.

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 43–67). Mahwah, NJ: Erlbaum.

Healy, J. (1998). *Failure to connect*. New York: Simon & Schuster.

Hofferth, S. L., & Sandberg, J. F. (2001a). How American children spend their time. *Journal of Marriage and the Family, 63*, 295–308.

Hofferth, S. L., & Sandberg, J. F. (2001b). Changes in American children's time, 1981-1997. In S. Hofferth & T. Owens (Eds.), *Children at the millennium: Where did we come from, where are we going?* (pp. 193–229). New York: Elsevier Science.

Jordan, A. (2004). The role of media in children's development: An ecological perspective. *Journal of Developmental & Behavioral Pediatrics, 25*(3), 196–206.

Lee, K. M., & Peng, W. (2006). What do we know about social and psychological effects of computer games? A comprehensive review of the current literature. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 327–345). Mahweh, NJ: Erlbaum.

Lieberman, D. A. (2006). What can we learn from playing interactive games? In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 379–397). Mahwah, NJ: Erlbaum.

McDonald, J., & Moffitt, R. (1980, May). The uses of tobit analysis. *Review of Economics and Statistics, 62*, 318–321.

Mitchell, E. (1985). The dynamics of family interaction around home video games. *Marriage and Family Review, 8*, 121–135.

Neuman, S. B. (1995). *Literacy in the television age*. Norwood, NJ: Ablex.

Nikken, P., & Jansz, J. (2006). Parental mediation of children's videogame playing: A comparison of the reports by parents and children. *Learning, Media and Technology, 31*(2), 181–202.

Peterson, J., & Zill, N. (1986). Marital disruption and behavior problems in children. *Journal of Marriage and the Family, 48*, 295–307.

Rathbun, A., West, J., & Hausken, E. G. (2003). *Young children's access to computers in the home and at school in 1999 and 2000* (Tech. Rep. No. NCES 2003-036). Washington, DC: U.S. Department of Education.

Roberts, D. F., Foehr, U. G., & Rideout, V. (2005). *Generation M: Media in the lives of 8-18 year-olds*. Menlo Park, CA: Henry J. Kaiser Family Foundation.

Roberts, D. F., Foehr, U. G., Rideout, V. J., & Brodie, M. (1999). *Kids & media @ the new millennium*. Menlo Park, CA: Henry J. Kaiser Family Foundation.

Rocheleau, B. (1995). Computer use by school-age children: Trends, patterns, and predictors. *Journal of Educational Computing Research, 1*, 1–17.

Rogers, S. J., Parcel, T. L., & Menaghan, E. G. (1991). The effects of maternal working conditions and mastery on child behavior problems. *Journal of Health and Social Behavior*, *32*, 145–164.

Rubin, A. (2002). The uses-and-gratifications perspective of media effects. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 525–548). Mahwah, NJ: Erlbaum.

Schmitt, K., & Anderson, D. (2002). Television and reality: Toddlers' use of visual information from video to guide behavior (Vol. 4, pp. 51–76).

Sefton-Green, J. (1998). Being young in the digital age. In J. Sefton-Green (Ed.), *Digital diversions: Youth culture in the age of multimedia* (pp. 1–20). London: UCL Press.

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Sotirovic, M. (2005). [Review of the book *Television and child development*, 3rd ed.]. *Journalism & Mass Communication Quarterly*, *82*, 477–478.

Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology*, *15*, 13–32.

Subrahmanyam, K., Greenfield, P., Kraut, R., & Gross, E. (2001). The impact of computer use on children's and adolescents' development. *Applied Developmental Psychology*, *22*, 7–30.

Timmer, S., Eccles, J., & Obrien, K. (1985). How children use time. In F. Juster & F. Stafford (Eds.), *Time, goods, and well-being* (pp. 353–382). Ann Arbor, MI: Institute for Social Research.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*, 24–36.

Van Evra, J. (2004). *Television and child development*. Mahwah, NJ: Erlbaum.

Vandewater, E., Bickham, D. S., & Lee, J. H. (2006). Time well spent? Relating television use to children's free-time activities. *Pediatrics*, *117*, 181–191.

Vered, K. O. (1998). Blue group boys play Incredible Machine, girls play hopscotch: Social discourse and gendered play at the computer. In J. Sefton-Green (Ed.), *Digital diversions: Youth culture in the age of multimedia* (pp. 43–61). London: UCL Press.

von Salisch, M., Oppl, C., & Kristen, A. (2006). What attracts children? In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 147–163). Mahwah, NJ: Erlbaum.

Weber, R., Ritterfeld, U., & Mathiak, K. (2006). Does playing violent video games induce aggression? Empirical evidence of a functional magnetic resonance imaging study. *Media Psychology*, *8*, 39–60.

Woodcock, R., & Mather, N. (1989). *W-J-R Tests of Achievement: Examiner's manual*. Allen, TX: DLM Teaching Resources.

Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Pinon, M., Scantlin, R., et al. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families: The Early Window Project. *Child Development*, *72*, 1347–1366.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1620–1631

Contagious Yawning in Autistic and Typical Development

Molly S. Helt and Inge-Marie Eigsti
University of Connecticut

Peter J. Snyder
Warren Alpert Medical School of Brown University

Deborah A. Fein
University of Connecticut

The authors tested susceptibility to contagious yawning in 120 children, 1–6 years, to identify the time course of its emergence during development. Results indicated a substantial increase in the frequency of contagious yawning at 4 years. In a second study, the authors examined contagious yawning in 28 children with autism spectrum disorders (ASD), 6–15 years. Children with ASD showed diminished susceptibility to contagious yawning compared with 2 control groups matched for mental and chronological age, respectively. In addition, children diagnosed with Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS) a milder variant of autism, were more susceptible to contagious yawning than were children diagnosed with full Autistic Disorder. The authors explore the implications of these findings for theories about the development of mimicry and emotional contagion.

Contagion refers to the tendency of a particular behavior to spread through a group in a "chain reaction." The behaviors that most often trigger contagious reactions in others are those that signify the inner states of others (Hatfield, Caccioppo, & Rapson, 1994). For example, infants in hospital nurseries begin to cry when they hear other babies crying (Hoffman, 1978; Simner, 1971), and laugh tracks accompany most television situational comedies because hearing the laughter of others prompts our own laughter (Bush, Barr, McHugo, & Lanzetta, 1989; Provine, 2000). Similarly, seeing another person yawn, thinking about yawning, reading the word *yawn*, or even hearing the word can elicit a yawn in 40%–60% of normal adults when exposed to such stimuli under experimental conditions (Baenninger & Greco, 1991; Platek, Critton, Myers, & Gallup, 2003; Provine, 1989). Such forms of behavioral contagion may reflect or facilitate *emotional contagion*, which refers to the tendency of individuals to converge emotionally with those around us. The property of contagion, as applied to emotions, offers psychologists an opportunity to study the roots of automatic social behaviors that potentially lay the foundations for the development of empathy.

At least one mechanism by which emotional contagion comes about is via the underlying processes of *mimicry* and *afferent feedback* (Hatfield et al., 1994). Whereas imitation involves the conscious, effortful reproduction of another's behavior, mimicry refers to nonvolitional "matching" behavior (Want & Harris, 2002). As we interact with someone, we often subconsciously mimic (often at a level undetectable to the naked eye or ear) their facial expressions, bodily postures, and speech patterns (see Niedenthal, Barsalou, Ric, & Krauth-Gruber, 2005, for a review).

When a person mimics, the activation of his or her emotional body schemas also creates an emotional reaction that corresponds to the movements being mimicked (i.e., the act of smiling causes us to feel happier). This was first known as *facial* feedback (Capella, 1993), but research has now documented similar effects for gesture, posture, and vocal prosody (see McIntosh, 1996, for a review), and so may be more accurately referred to as *afferent feedback* (Hatfield et al., 1994). The coupling of our automatic tendency to mimic others and the effects of afferent feedback on our own emotional

---

These data were collected as part of the requirement for completion of the masters thesis of the first author. We thank Kastley Marrin and Melissa Fernandez for their patience and attention to detail in coding and entering data.

Correspondence concerning this article should be addressed to Molly S. Helt, Department of Psychology, University of Connecticut, 406 Babbidge Rd, Unit 1020, Storrs, CT 06269. Electronic mail may be sent to molly.helt@uconn.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8105-0021

states may explain the ubiquity of emotional contagion: Because we unconsciously match the emotional movements of others, we unconsciously feel the emotions of others as we interact with them. Similarly, this account of mimicry may explain its psychological utility: Emotional contagion may allow human beings to intuit the feelings of those around us when we interact socially (Rogers & Williams, 2006).

Yawning is a type of mimicry in that it is a matching behavior that is produced unconsciously. The robust effects for contagious yawning found by Provine et al. (1986) when they exposed participants to videos of people yawning disappeared when the participants knew they were being watched (Provine, 2005), supporting the idea that contagious yawning is an unconscious process that may actually be suppressed by conscious awareness. However, contagious yawning appears to be a special kind of mimicry. Unlike the miniscule muscular movements, often undetectable to the naked eye, that mimicry typically entails, yawns—large, obvious sequences of movements, difficult to stop once initiated and lasting for up to 10 s—are a very visible result of such mimicry. The explanation for this may derive from the fact that yawning is a *fixed action pattern* (Provine, 1986). A fixed action pattern is a species-typical behavioral sequence that is indivisible and, once initiated, runs to completion. Fixed action patterns are invariant and are triggered by a neural network called the "innate releasing mechanism" in response to a "sign stimulus" or "releaser" (in this case, the yawns of others act as an innate releasing mechanism; Tinbergen, 1951). Perhaps the phenomenon of *contagious* yawning can be explained by the idea that mimicking the first part of a fixed action pattern (as would be expected if one were mimicking the facial expressions or vocalizations of a conspecific as he or she begins to yawn) is likely to trigger the release of the entire behavior.

In addition to being a form of mimicry, contagious yawning appears to involve an emotional component. Deputte (1994) identified two contexts for yawns: the "rest yawn" observed when the context involves a change in arousal level, and the emotion yawn (which he noted could also be referred to as a "social yawn"). He defined the emotion yawn as an action used as an unconscious communication of psychological decompression after a state of high alert. Yawning is also similar to the other highly contagious acts of crying and laughing, both of which represent emotional states, in that it produces a distinct sound, as well as a distinct facial expression (Provine, 1996). The existence of an emotion yawn suggests that contagious yawning may be considered a form of emotional contagion. On the other hand, yawning may not signal an emotion but may simply be a facial expression that is unintentionally mimicked, as are other nonemotional facial expressions (Heyes, 2001). If so, then its contagiousness may simply be a by-product of ubiquitous facial mimicry, perhaps evolutionarily adaptive because it facilitates contagion of "true" emotions. In either case, the disruption of mimicry, which may be demonstrated by a disruption in contagious yawning, should have consequences for emotional resonance with others. Indeed, empathetic people exhibit greater amounts of all forms of mimicry (Chartrand & Bargh, 1999) and are reportedly more susceptible to contagious yawning (Platek et al., 2003).

Development of Mimicry, Emotional Contagion, and Contagious Yawning

Some capacity for mimicry and emotional contagion appears to be present from birth. Infants as young as 3 hr old execute some behaviors after witnessing them modeled by an adult (Meltzoff & Moore, 1977), most reliably, tongue protrusion (for a review, see Rogers, 2006). Similarly, a case has been made for the presence of emotional contagion from the first days of life; specifically, newborn infants will cry when they hear other babies crying in the nursery, but not in response to computer-generated sounds matched for acoustic properties (Simner, 1971). According to one point of view, newborn "matching" behaviors are early examples of the strong link between the *perception* of actions and emotions in others, and the *experience* of these actions and emotions in ourselves, thus setting the stage for imitative abilities that emerge later and underlie critical aspects of social and emotional development. However, early mimetic behaviors and contagious crying gradually become less frequent after birth. Imitative tongue protrusion disappears by 2–3 months (Abravanel & Sigafoos, 1984), and whereas 84% of newborns exhibit contagious crying, only 24% of 3- to 12-month-old infants do so (Bühler & Hetzer, 1928). In place of full-blown contagious crying, infants at 10–14 months are much more likely to mimic the distressed facial expressions of other crying children (Zahn-Waxler, Radke-Yarrow, & King, 1979). Given these decreases in matching behaviors after 2–3 months, it is unclear whether these early behaviors are continuous with later developing capacities for mimicry and emotional contagion. Mimicry and emotional contagion seem to ''reappear'' during the second half of the 1st year and then continue to expand thereafter. For example, by 9 months, babies take on the basic moods and facial expressions of their caregivers, such as joy and sadness (Termine & Izard, 1988). A longitudinal study by Jones (1996) found that mimicry of distinct actions appeared at distinct periods in development between 6 and 18 months.

Spontaneous yawning (termed the ''rest yawn'' according to Deputte, 1994) begins in utero. In contrast, the limited data on contagious yawning (presumably, the emotion yawn) hint that it may not emerge until much later in development. Piaget (1951) reported that his own children did not begin to yawn contagiously until the 2nd year of life. A more recent study suggested an even later age of emergence for contagious yawning, reporting that only children 5 years and older yawned in response to videos of yawns (Anderson & Meno, 2003). Children respond to live stimuli well before they respond to video stimuli (Troseth & deLoache, 1998), and so in vivo yawning may be a more effective stimulus for eliciting contagious yawning in younger children. Furthermore, the instructions in the Anderson and Meno (2003) study required participants to clap whenever they witnessed a yawn. Clapping is an arousing activity, and increased arousal is associated with diminished yawning (Provine, 2005). In addition, the clapping instructions drew conscious attention to yawning, which has also been shown to diminish yawning (Provine, 2005). For all these reasons, the extant results must be interpreted with caution. Therefore, there are few data to address the developmental course of contagious yawning in typical development.

Mimicry, Emotional Contagion, and Contagious Yawning Deficits in Individuals With ASD

Consistent with the hypothesis that mimicry facilitates emotional contagion (Hatfield et al., 1994), children with autism spectrum disorders (ASD) show deficits in both. McIntosh, Reichmann-Decker, Winkielman, and Wilbarger (2006) found that, unlike controls, adults with high-functioning autism did not unconsciously copy the facial expressions of individuals viewed on a video screen (despite being able to imitate them when asked). Meanwhile, children with ASD have been found to be less susceptible to emotional contagion when provided with prompts that typically elicit emotions that match those of the model: for example, an experimenter opening a box in a child's presence and looking delighted or afraid (Scambler et al., 2006) or an experimenter injuring herself (Bacon, Fein, Morris, & Waterhouse, 1998).

One study has examined contagious yawning in children and adolescents with diagnoses including Autistic Disorder, Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS), and Asperger's disorder. Participants with ASD were less susceptible than age-matched typically developing (TD) peers to contagious yawning when exposed to video recordings of people yawning (Senju et al., 2007). Results are difficult to interpret, however, because the ASD group had a lower mental age, and included fewer females, and because of the use of video rather than live models. Given that all these factors could potentially play a role in mimicry, further data are needed to clarify the dimensions of the relative susceptibility of individuals with ASD to contagious yawning.

Current Studies

Study 1 explored the chronological and mental ages at which contagious yawning is exhibited in typical development to provide clues about psychological mechanisms underlying this phenomenon. The study delivered yawning stimuli in an implicit paradigm (i.e., no instructions were given regarding yawning) by a live model, and study procedures were the same for all children—design factors intended to maximize sensitivity to the phenomenon. If contagious yawning is a primitive mimetic response, like contagious crying, it should be evident in the youngest members of the sample. In this case, its reduction or absence in individuals with ASD would imply that basic building blocks of social connection may be abnormal from birth. In contrast, a protracted emergence of contagious yawning would be consistent with the idea that mimicry becomes more ubiquitous and/or fine-tuned as development proceeds. In this case, diminished yawning contagion among individuals with ASD would seem to be a reflection of their diminished implicit social and emotional learning and experience.

Study 2 sought to measure the rates of contagious yawning in a group of children with ASD. This study provided more rigorous diagnosis than Senju et al. (2007), via use of the Autism Diagnostic Observation Scale (ADOS; Lord et al., 2000), excluding children with Asperger's disorder (a variant of the disorder in which language is not affected and potentially involving distinct neural systems; Goodman, 2005), and using live, rather than videotaped, stimuli. In addition, the study sought to examine whether these group differences could be accounted for by differences in mental age, gender, symptom profile, or diagnosis.

Study 1

Method

Participants

Participants were 123 TD children, ages 1–6 years (approximately 20 at each year of age studied), recruited through two local day-care centers. Parents, teachers, and day-care workers were asked to report any known or suspected developmental difficulties for each child. For children age 2 and above, parents or teachers completed the Behavioral Assessment System for Children, Second Edition (BASC–2; Reynolds & Kamphaus, 2004) to screen for any developmental problems. The data of 3 children were excluded—one due to concerns about Tourette's syndrome, one due to coding problems, and one due to lack of attention —for a final sample of 120 children. Sex distribution varied for each year of age with approximately 57% ($n = 68$) of the total sample being female (see Table 1 for gender breakdown by age). The ethnic background of this group was as follows: 3 children were African American, 5 children were Asian, 32 children were Latino, 3 children were multiethnic, and 80 children were Caucasian.

Procedure

A letter describing the study and obtaining consent was sent home to parents. Parents were asked to either fill out a BASC–2 for their child or give permission for the child's teacher to fill out a BASC–2 for the child. Children whose parents consented to the study were taken individually from their classrooms into a quiet room, seated across from the experimenter. After introducing herself and explaining, ''Now we're going to read some stories together, and then you can pick a prize from this basket'' and showing the child the various toy prizes, the experimenter read aloud one to four stories (depending on the child's age and the complexity of the stories) for a total reading time of approximately 12 min (sometimes with a brief break, again dependent upon the child's developmental level). During the first 2 min of reading, the experimenter did not yawn at all. This, in addition to the first few minutes of introductions before reading, was to control for any changes in arousal level that may have occurred when the children sat down to listen to a story. During the last 10 min of reading, the experimenter paused four times to yawn and discreetly recorded on a coding sheet when a child yawned. Approximately 40% of the sessions (49 of 120), randomly selected, were videotaped and coded by two independent raters (who had not assisted in data collection) for reliability.

For coding purposes, contagious yawns were defined as yawns occurring within 90 s of the yawning stimulus. Although subconscious reactions typically occur much more rapidly, previous reports of contagious yawning (e.g., Provine, 2005) indicate that contagious yawning reactions may be somewhat slower and we wanted to allow for the most liberal possible estimate of the phenomenon, as well as for the possibility that yawning contagion may be mediated by some cognitive processes. Because contagious yawning is a low base rate phenomenon, yawning was coded as a dichotomous

Table 1
Characteristics of Sample in Study 1

	1–2 years ($n = 20$)	2–3 years ($n = 20$)	3–4 years ($n = 20$)	4–5 years ($n = 20$)	5–6 years ($n = 20$)	6–7 years ($n = 20$)
Chronological age (months)						
M	18.1	29.6	41.7	53.2	65.8	79
SD	3.1	3.2	3.3	3.2	3.3	3.5
Range	13–23	24–35	36–47	48–58	61–71	72–83
Visual attention scores (possible range 1–16)						
M	13.2	13.7	14.25	14.65	15.1	14.6
SD	1.4	1.26	1.3	1.01	0.91	0.99
Range	12–16	12–16	12–16	12–16	13–16	13–16
Gender (% female)	60%	55%	55%	55%	60%	55%

variable (present or absent) rather than a measure of yawning frequency (number of yawns total). (Analysis of yawning frequency in all statistics examined in Studies 1 and 2 did not yield a different pattern of results.) Children were considered to be contagious yawners if they yawned in response to at least one of the experimenter's four yawns within a 90-s window. (Yawns occurring outside the 90-s window were exceedingly rare, with only two occurring over the 196 videotaped trials.) Coding criteria required the presence of all of the physical manifestations of a yawn (Provine, 2005): open mouth, narrowed eyes, and an indrawn breath. The data of one 3-year-old girl had to be excluded because she covered her mouth and made a sighing sound but her eyes did not narrow at all. Thus, neither the experimenter nor the research assistants were able to determine whether the child was pretending to yawn in order to imitate the experimenter or was truly yawning. Interrater reliability was 100%, when the one ambiguous responder was removed.

Videotaped sessions were also coded for children's looking behavior toward the experimenter during the 10-s segment in which the yawns were presented by assigning each child a rating of 1 *(not looking at all)*, 2 *(looking for less than 5 s of the 10 s block coded)*, 3 *(looking for 5 or more seconds of the 10 s block coded)*, or 4 *(looking throughout the 10 s block in which the yawn stimulus was presented)* for each of the four trials, resulting in a visual attention score of 1–16 for each child. The data of children with scores of less than 12 ($n = 1$ in Study 1) were excluded from analysis.

Results and Discussion

Results showed that, when exposed to the yawns of a live model over four trials, the proportions of children in each age group who yawned contagiously for at least one trial were as follows: 0% (0/20) of 1-year-olds, 5% (1/20) of 2-year-olds, 10% (2/20) of 3-year-olds, 35% (7/20) of 4-year-olds, and 40% (8/20) of both 5- and 6-year-olds (see Figure 1). Binary logistic regression revealed a significant effect for age, $\chi^2(1, N = 120) = 22.254$, $p < .001$, but not for visual attention, $p = .22$, or gender, $p = .37$, in predicting whether or not a child will yawn contagiously at least once across the four trials of the experiment, with each year of age increasing the odds of contagious yawning by a factor of 2.2. However, the data suggest that these variables do not have a linear relation. Thus, age was recoded dichotomously as “under 4 years” or “4 years and over,” $\chi^2(1, N = 120) = 22.69$, $p < .001$, revealing that being 4 years or older increased each child's odds of yawning contagiously by a factor of 14.7 (visual attention and gender remained insignificant).

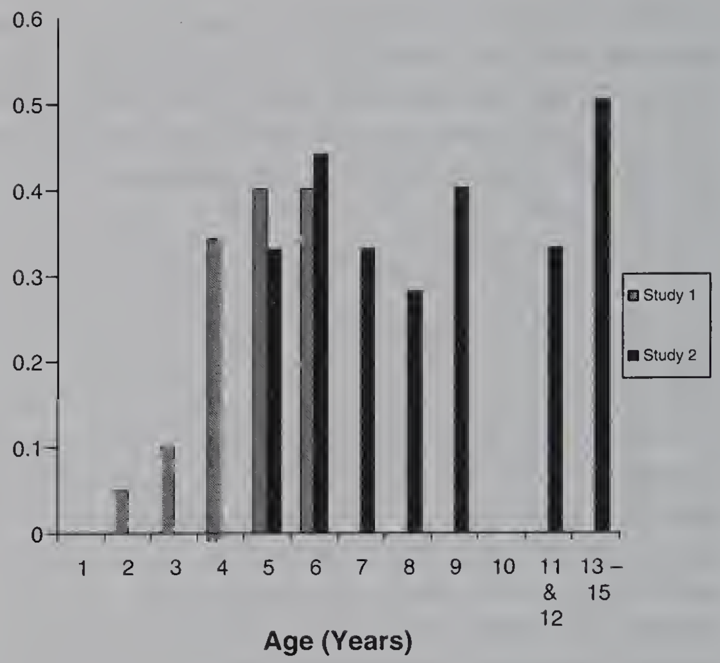


Figure 1. Proportion of contagious yawning by chronological age in typically developing children, Study 1 (children 1–6 years) and Study 2 (children 5–15 years).

Although the older children appeared to make more sustained eye contact in the direction of the experimenter's face and the book throughout the session, the experimenter yawned when she had each child's attention, resulting in good visual attention to the yawns from all age groups (see Table 1). The amount of visual attention that children directed toward the examiner's face was examined in one-way analysis of variance. There was a statistically significant difference in visual attention scores for the six age groups, $p < .01$. Post hoc comparisons indicated that the mean attention score for 1-year-olds was significantly different from that for 5-year-olds, $p < .001$, and 6-year-olds, $p = .02$. In addition, the mean score for the 2-year-olds differed from that for the 5-year-olds, $p < .01$.

These results indicate that children younger than 4 years are significantly less likely than children 4 years and older to yawn after being exposed to the yawn of another. In contrast to Anderson and Meno (2003), who observed no contagious yawning prior to age 5, the present study observed contagious yawning as early as 2 years with a significant shift in frequency at 4 years. The differing results between the present data and those of Anderson and Meno (2003) may be the result of presenting the yawning stimuli in a live, implicit paradigm rather than a video, explicit paradigm. Only a single child exhibited contagious yawning at age 2—the age at which Piaget (1951) observed his children to begin yawning contagiously. Given that individuals tend to engage in more mimicry with individuals with whom they have rapport (Chartrand & Bargh, 1999), parents may elicit contagious yawning in their own children earlier than it can be measured by an unfamiliar experimenter. It is also possible that parents may notice a pattern of earlier onset because they have a much larger window in which to observe their children compared with this relatively brief experiment.

The temporal discrepancy between the emergence of spontaneous (rest) yawning and contagious crying (evident from birth), and that of contagious (social-emotional) yawning, indicate that they may not share a common basis. Rather, this later emergence of contagious yawning suggests that it may build upon early empathic or cognitive development.

Study 2

Method

Participants

Participants were 30 children with a diagnosis of an ASD, in addition to 63 TD children. Diagnosis for the ASD group was confirmed by the experimenter (MH) using the Autism Diagnostic Observation Scale (ADOS) using *DSM–IV* criteria (American Psychiatric Association, 1994); see Table 2 for group characteristics. Fifteen children met criteria for autistic disorder on the ADOS, whereas 13 children met criteria for a milder form of autism (PDD-NOS). PDD-NOS refers to cases in which a child experiences marked impairment in social interaction, as well as difficulties with communication and/or stereotyped behaviors patterns or interests but does not meet full diagnostic criteria for autistic disorder. The ethnic background of this group was as follows: One child was African American, 1 child was Asian, and 28 children were Caucasian.

The two TD groups were matched to children with ASD on either chronological age (within 6 months) or mental age (within 6 months) as well as gender. No TD child was used in both mental age (MA) and chronological age (CA) control groups, and none of the children in Study 2 participated in Study 1. The data of 7 TD children were excluded because they were not matched to the ASD group on these parameters. Children with ASD were recruited via flyers at local autism events whereas TD children were recruited via flyers at local family events. In the TD group, 2 children were Asian, 12 children were Latino, 3 children were multiethnic, and 46 children were Caucasian.

Measures

The Socialization domain of the Vineland Scales of Adaptive Behavior–Interview Edition. The Vineland (Sparrow, Balla, & Cicchetti, 1984) is a semistructured parent interview designed to assess a child's adaptive functioning. This study utilized the Socialization domain. Age-equivalent and standard scores for social skills were used in the present analyses.

Autism Diagnostic Observation Scale. The ADOS (Lord et al., 2000) consists of a structured play session that provides participants opportunities to engage in conversation, narrative, problem-solving,

Table 2
Characteristics of Autism Spectrum Disorder (ASD) and Typically Developing (TD) Control Groups in Study 2

	ASD (*n* = 28)	Mental age-matched TD (*n* = 28)	Chronological age-matched TD (*n* = 28)
Chronological age	10:5 (2.2); 6–15 years	9:1 (2.1); 5–13 years	10:3 (2.3); 6–15 years
Mental age (Stanford–Binet)	9:4 (2.0); 5:8–13:4 years	9:2 (2.2); 5:6–13:6 years	12.5 (1.9); 6:2–16:10 years
Vineland (standard) socialization scores			
M	69	101	103
SD	15.87	10.67	11.1
Range	44–100	87–115	92–121
Gender	93% (*n* = 26) male	93% (*n* = 26) male	93% (*n* = 26) male

Note. Results given as mean (standard deviation); range.

and imaginative tasks. The ADOS is scored according to *DSM–IV* criteria for the autism spectrum diagnoses.

Stanford–Binet Intelligence Scale: Fifth Edition (Roid, 2003). In the Stanford–Binet Abbreviated IQ scales, children provide word definitions, yielding a verbal IQ score, and solve a series of picture puzzles, yielding a performance IQ score. This measure provided an index of MA.

Procedure

Children with ASD and TD children sat across the table from a live examiner. Children were tested in a quiet room in their home or in the University of Connecticut Psychological Services Clinic. After an introduction and discussion of the procedure, the experimenter read each child a story, yawning four times during the story. She then administered the Stanford–Binet and (for the ASD group) the ADOS. Because older children, and those who are fluent readers, may have felt uncomfortable with an adult reading to them, those children were told, ''First, I'm going to read you a story, and then I'll ask you some questions about it'' as opposed to simply, ''First, I'm going to read you a story.'' Thus, 14 of the 28 participants in both the ASD and mental-age-matched typically developing control (MA-TD) groups and 22 of the 28 participants in the chronological-age-matched typically developing control (CA-TD) group were told they would be asked questions following the story. In order to explore whether children with ASD yawned spontaneously (as opposed to contagiously) less often than TD children, the test period of the Stanford–Binet administered after the yawning stimuli were also coded for the presence or absence of yawns.

All story sessions were recorded on videotape and coded for looking and yawning behavior by raters blind to group status using the same coding procedures as in Study 1. Interrater reliability, coded for each block, was 83% for looking behavior and 100% for yawning. After the story and test session, parents completed the socialization portion of the Vineland interview.

Results and Discussion

Children with ASD were significantly less likely than TD children to yawn after being exposed to the experimenter's yawn. Only 11% (3/28) of the children with ASD aged 5–12 yawned, compared with 43% (12/28) of a TD control group matched for CA-TD, Fisher's exact $p = .009$, and 36% (10/28) of a control group matched for MA-TD, Fisher's exact $p = .037$ (see Table 3 for group characteristics). When the ASD group was further divided by diagnosis, it became clear that children diagnosed with autistic disorder were less likely to exhibit contagious yawning than children diagnosed with PDD-NOS, $\chi^2(1) = 3.88$, $p = .05$. Specifically, *none* of the children with autistic disorder ($n = 15$) showed contagious yawning, whereas 23% of the children with PDD-NOS ($n = 13$) did. CA, MA, instruction type, visual attention scores, Vineland socialization scores, and diagnosis were entered into a single binary logistic regression model with contagious yawning as a dichotomous variable. In contrast to the results for younger children (ages 1–6 years) found in Study 1, binary logistic regression on contagious yawning in children ages 5–15 years revealed no main effect for age, $p = .80$ (see Figure 1). These data are consistent with the possibility that the rate of contagious yawning remains relatively stable after 5 years, though the relatively small sample size for the current study prevents us from drawing a firm conclusion. No significant effect was found for instruction type ($p = .83$) or MA ($p = .87$). Surprisingly, no significant effect was found for Vineland socialization scores ($p = .924$). The only significant predictor of contagious yawning was diagnosis ($p = .014$), with the absence of an autism spectrum diagnosis increasing the odds of contagious yawning by a factor of 5.8.

A second binary logistic regression was used on the ASD group data only to determine whether autism symptom severity (as indexed by ADOS scores) showed a relation with contagious yawning. Despite the effect of diagnostic severity on yawning contagion, the results suggested no relation between contagious yawning and total ADOS scores, $\chi^2(1, N = 28) = 1.08$, $p = .30$, or communications ($p = .276$) and socialization ($p = .919$) subscales. A possible explanation for this discrepancy is that although the ADOS is an exceptional

Table 3
Number of Children Who Exhibited Contagious Yawning in Study 2 by Group

	Autistic disorder	PDD-NOS	MA-TD	CA-TD
Yawn	0	3	10	12
No yawn	15	10	18	16
% total	0%	23%	36%	43%

Note. PDD-NOS = Pervasive Developmental Disorder, Not Otherwise Specified; MA-TD = mental-age-matched typically developing control; CA-TD = chronological-age-matched typically developing control.

instrument for diagnosing autism, ADOS scores have not been standardized for the purposes of using it as a continuous measure of autism severity. Indeed, individual items on the ADOS vary with respect to their correlation with both IQ and CA (Gotham, Pickles, & Lord, 2009). It is also likely that the very low rate of contagious yawning in the ASD group may have masked relations between contagious yawning and symptom profiles in this group.

Two children with ASD were excluded from analysis because their visual attention scores fell below 12. For the remaining children, ASD ($M = 14.6$) and TD ($M = 15.1$) groups did not differ significantly, $t = .74$, $p = .14$, on the amount of time spent looking toward the experimenter during the 10 s periods in which the yawning stimuli were presented. As in Study 1, this is likely because the experimenter yawned at times when she was confident that she had the child's attention, not because there are no general differences in visual attention between groups.

Finally, TD children were no more likely to yawn (19.6%) than children with ASD (14.3%) during the nonexperimental portion of the study (such as the Stanford–Binet administration), $\chi^2(1, N = 28) = .091$, $p = .76$. In addition, the experimenter asked the parents of the ASD children whether they believed that their child yawns less frequently than TD children; only one parent responded in the affirmative. Taken together, these data imply that the ASD deficit observed is specific to the contagious, or emotion yawn and does not affect the spontaneous, or rest yawn.

The diminished tendency for participants in the ASD group to yawn contagiously to a live stimulus is consistent with previous findings that individuals with ASD show diminished contagious yawning with a video stimuli (Senju et al., 2007) as well as general deficits in mimicry (McIntosh et al., 2006). The current study is novel in several regards. First, the data demonstrate that individuals with ASD are less likely to yawn contagiously than their TD peers, even when controlling for MA and gender. Second, the magnitude of this difference in autistic disorder is much greater than previously reported, when using live rather than videotaped stimuli. Most importantly, the current data demonstrated a striking effect of diagnosis, with children with autistic disorder being less likely to show contagious yawning than children with PDD-NOS. Autistic disorder is characterized by more marked deficits in the realms of communication, social relatedness, and stereotyped interests and behaviors. Thus, the latter findings indicate a significant relation between diagnostic severity in ASD, and susceptibility to contagious yawning.

At a broad level, the late onset of this phenomenon in typical development implies that an autistic deficit in contagious yawning may be a reflection of a deficit in early social learning that affects the ability or tendency to mimic the actions and emotions of others. Nonemotional sounds and mouth movements, such as mouth opening when feeding a baby, are unintentionally imitated (Heyes, 2001) and modulated by experience (Heyes, 2005). Future research is needed to explore whether these forms of mimicry are unconsciously imitated by children with autism (work that is currently underway in this lab). If they are not, this finding would lend further support to the proposal that there is an autistic deficit in recognizing or acting on the correspondence between oneself and others (an ability that all mimetic and imitative acts depend upon). Alternatively, individuals with autism may demonstrate a specific deficit in mimicking emotional behavior. Indeed, emotion perception is a documented area of weakness for individuals with autism (e.g., Moody, McIntosh, Mann, & Weisser, 2007).

General Discussion

The present study demonstrates that children younger than 4 years are significantly less likely to yawn contagiously when exposed to a live stimulus—that is, after seeing another, nearby person yawn—than children ages 4 years and older. The very early emergence of *spontaneous* yawning (by the end of first trimester of pregnancy) offers an informative contrast to the much later, postnatal development of *contagious* yawning. This developmental lag may reflect the phylogenetic antiquity of the motor act of yawning (demonstrated in most vertebrates) relative to the recent evolution of contagion (convincingly demonstrated only in humans and chimpanzees; Anderson, Myowa-Yamakoshi, & Matsuzawa, 2004). The present study also demonstrates that children with ASD are significantly less likely to yawn contagiously when exposed to a live stimulus than TD children. Furthermore, those with a diagnosis indicative of more severe autistic symptoms (autistic disorder) are significantly less likely to yawn contagiously than those with a diagnosis indicative of milder autistic symptoms (PDD-NOS). Taken together, these studies provide further validation for the distinction between spontaneous (or ''rest'') yawns and contagious (or ''emotion'') yawns, and suggest that contagious yawning is linked with social development.

More research is needed to establish whether contagious yawning is unique or whether all forms of mimicry increase over the course of typical development (potentially with a substantial shift in competence occurring around the age of 4). However, one potential implication of the late onset of contagious yawning in typical development and its near absence among individuals with ASD is that mimicry, and hence emotional contagion, may increase with social experience. It is fairly easy to imagine how mimicry may come to substantially increase during the first years of life. First, it is adaptive for an individual to learn to ''automatically'' experience the emotions displayed on the faces of conspecifics (i.e., experience those emotions without deliberation or analysis). For example, if an individual becomes afraid and retreats upon seeing a predator, it is adaptive for an individual to retreat rather than analyze the situation. Likewise, automatically experiencing a conspecific's disgust may protect an individual from ingesting the same harmful food item (Goldman, 2005).

In the case of yawning, experiencing fatigue along with conspecifics may have served the primary evolutionary function of synchronizing a social group's biological rhythms (Schurmann et al., 2004). Indeed, there is some evidence that social cues alone may sustain shared circadian rhythms in the absence of light (Aschoff et al., 1971). Alternatively, rather than serving a direct evolutionary purpose, contagious yawning may have arisen as a by-product of the automatic tendency to facially mimic others (adaptive in rapidly spreading signals of risk and reward throughout a group).

In addition to facilitating emotional and biological synchronization, mimicry may serve to increase feelings of closeness and connection between individuals (Lakin, Jefferis, Cheng, & Chartrand, 2003). These feelings of affiliation, in turn, increase the amount of mimicry individuals display toward one another (Lakin et al., 2003). This process may escalate in a continuous cycle first between child and parent and, later, child and other members of the community, during the early years of development. Out of such increasingly synchronous bonds may develop abilities such as speech (i.e., Bruner's ''proto-conversations''; Bruner, 1983), joint attention, and deliberate imitation.

It is possible that infants with autism are not able to perceive that they are being mimicked (due to reduced social attention or difficulties with emotion perception) or are unable to mimic accurately (due to difficulties in motor function or self-other mapping), and so synchronization with mother and other members of the social group is less available throughout development. This lack of early mimicry could also affect feelings of psychological connection and opportunities for social learning. These changes could thus leave children with autism unable to recognize primitive socioemotional cues that could otherwise serve to biologically and emotionally synchronize them with people around them.

Speculation About Underlying Neural Mechanisms

Brain imaging studies may hold some promise for shedding light on the neurological substrate of contagious yawning. Two neuroimaging studies have explored contagious yawning by exposing participants to videos of others yawning. One study (Platek, Mohamed, & Gallup, 2005) showed increased activation in the precuneus and posterior cingulate gyrus when participants viewed videos of yawning relative to videos of laughter (another emotionally contagious stimulus). The authors interpreted this activation as evidence that contagious yawning, compared with laughter, involves nonconscious aspects of self-referencing. A second study (Schurmann et al., 2004) found that watching yawn videos elicited activation in a region associated with perception of biological movement—particularly perception of eye and mouth movements (the superior temporal sulcus). In addition, they reported that self-reported desire to yawn was negatively correlated with periamygdalar activation (a region implicated in arousal)—a finding that likely derives from the negative relation between arousal and yawning. Neither study found any special involvement of the *mirror neuron system* (MNS)—a system that is active both when one observes and executes an action—for yawning over and above that recruited for the control stimuli (laughing or neutral faces, and meaningless mouth movements, respectively). An obvious limitation of these studies is that the participants in these studies did not ''catch'' the yawns; they simply reported upon their desire to yawn while viewing the yawning stimuli and thus did not actually experience any form of emotional contagion.

Studies investigating forms of empathy and emotional convergence that more readily lend themselves to neuroimaging have consistently found activation in the insula and anterior cingulate cortex (ACC). We speculate that these areas are likely candidates for distinguishing the neural activation associated with contagious yawning over and above that associated with any type of action (and thus the MNS) including spontaneous yawning or other mouth movements, that is, the regions we speculate underlie the *contagious* component of yawning. Just as the "MNS" may underlie common coding of actions, studies have demonstrated the same common coding in the insula and ACC between first- and third-person experiences of emotional information, such as disgust (Wicker et al., 2003), pain (Singer et al., 2004), emotional body language (de Gelder & Hadjikhani, 2006), and emotional expressions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). A recent study of children ages 9–10 years supports the involvement of the insula, as well as inferior frontal mirror areas and the amygdala, in the observation and imitation of emotional expressions (Pfeifer, Iacoboni, Mazziotta, & Dapretto, 2008). Furthermore, activation in these areas was associated with the children's empathy levels. In children with autism, activity in these same areas during the observation of emotional expressions has been found to be negatively correlated with symptom severity (Dapretto et al., 2006).

One type of neuron unique to the insula and the ACC is the spindle cell, also known as the Von Economo neuron (VEN). Spindle cells are a late-evolving class of neurons that allow for the conscious monitoring of visceral signals from the body, a process that may underlie the impact of afferent feedback. These neurons do not achieve their ultimate postnatal number until the age of 4 years (the age at which we demonstrate the robust emergence of contagious yawning) and have been reported to be disordered in autism (Allman, Watson, Tetrault, & Hakeem, 2005). VENs are responsible for rapidly extracting statistical probabilities from sensory input based on previous information, giving rise to nonconscious emotional reactions. The more experience we gain, the more we may be able to recognize such patterns and associations, and the more developed these neurons may become. The diminished tendency for contagious yawning in children with ASD could be the result of primary or secondary neurological differences affecting these neurons.

Limitations

While study results were fairly unambiguous, findings are necessarily limited in several dimensions. First, although looking time, or general attention to the experimenter, was accounted for in the present studies, we could not unequivocally verify whether participants were attending to the eye or the mouth region of the speaker; this would have required additional eye tracking methods. This limitation is relevant because Provine (1989) has found that yawning contagion, when elicited by pictures of others yawning, is dependent upon attending to the region of the eyes rather than the mouth. Eye tracking studies have revealed that individuals with autism preferentially fixate on the mouth rather than the eye regions (Klin, Jones, Schultz, Volkmar, & Cohen, 2002), leaving open the possibility that a lack of attention to the *eye region* of others' faces (rather then a deficit in unintentional mimicry) is responsible for diminished yawning contagion in this group. Similarly, as typical children age, they may pay closer attention to the faces of others, resulting in more facial mimicry, and hence, greater emotional and biological attunement with those around them. In contrast, children with autism may never learn to pay closer attention to the faces of those around them, and so their tendency to mimic others, as well as their emotional attunement with others, may not increase over time in the way one observes in typical development. If so, this lack of attention to faces may have a previously unsuspected significance in that it may serve to impair mimicry, resulting in decreased emotional resonance, and perhaps even decreased biological synchrony with others.

An argument against this explanation of the findings is that yawning, like crying and laughing, is a contagious act that produces a sound, and the sound alone is often sufficient to trigger its contagion (Provine, 1996). Indeed, the experimenter in this study produced yawns that included the prototypical auditory accompaniment of an indrawn breath followed by a voiced sigh. In addition, given that there were no significant differences in visual attention to the yawns between the 4-year-olds (the age at which a major susceptibility shift was observed) and the other age groups, the developmental results seem unlikely to be due merely to age-related shifts in visual attention.

Finally, although the low rate of contagious yawning in the ASD group speaks to its significance, it made it impossible to determine the extent to which contagious yawning in this population, when it does occur, is correlated with specific autistic symptoms. If more children with ASD could be found who do yawn contagiously, it might be possible to determine whether such factors correlate with contagious yawning in autism, shedding increased light on the implications of this deficit.

Summary and Conclusions

In order to elucidate the developmental properties of automatic facial mimicry in the form of contagious yawning, we investigated the extent to which groups of children at various stages of social development (TD children ages 1–6, as well as children with ASD) are susceptible to contagious yawning. Findings strongly suggested that children under the age of 4 and children with ASD are less likely to yawn when exposed to another's yawn. The developmental curve associated with the onset of contagious yawning implies that emotional contagion becomes more developed and more sensitive over time, resulting in increased affective attunement with others as children grow older. Meanwhile, individuals with ASD may not experience increased emotional contagion during the early years of development, causing them to be deficient in the automatic emotional reciprocity that psychologically binds most individuals together. More research is necessary to determine what mediates this phenomenon; however, it is possible that in the future contagious yawning may prove to provide a simple measure of automatic facial mimicry, empathy, or perhaps even a biomarker of clinically important neurological characteristics that does not require specialized equipment or testing to detect.

References

Abravanel, E., & Sigafoos, A. D. (1984). Exploring the presence of imitation during early infancy. *Child Development*, *55*, 381–392.

Allman, J. M., Watson, K. K., Tetrault, N. A., & Hakeem, A. Y. (2005). Intuition and autism: A possible role for Von Economo neurons. *Trends in Cognitive Neuroscience*, *9*, 367.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington, DC: Author.

Anderson, J. R., & Meno, P. (2003). Psychological influences on yawning in children. *Current Psychology Letters*, *11*. Retrieved from http://www.baillement.com/replication/anderson-children.html.

Anderson, J. R., Myowa-Yamakoshi, M., & Matsuzawa, T. (2004). Contagious yawning in chimpanzees. *Proceedings of the Royal Society of London: Biological Sciences*, *271*, S468–S470.

Aschoff, J., Fatranská, M., Giedke, H., Doerr, P., Stamm, D., & Wisser, H. (1971). Human circadian rhythms in continuous darkness: Entrainment by social cues. *Science*, *171*, 213–215.

Bacon, A. L., Fein, D., Morris, R., & Waterhouse, L. (1998). The responses of autistic children to the distress of others. *Journal of Autism and Developmental Disorders*, *28*, 129–142.

Baenninger, R., & Greco, M. (1991). Some antecedents and consequences of yawning. *The Psychological Record*, *41*, 453–460.

Bruner, J. S. (1983). *Child's talk: Learning to use language*. New York: Norton.

Bühler, C., & Hetzer, H. (1928). Das erste verstandnis fur ausdruck im ersten lebensjahr [The understanding of expression in the first year of life]. *Zeitschrift for Psychologie*, *107*, 50–61.

Bush, L. K., Barr, C. L., McHugo, G. J., & Lanzetta, J. T. (1989). The effects of facial control and facial mimicry on subjective reactions to comedy routines. *Motivation and Emotion*, *13*, 31–52.

Capella, J. N. (1993). The facial feedback hypothesis in human interaction. *Journal of Language and Social Psychology*, *12*, 13–29.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, USA*, *100*, 5497–5502.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Sigman, M., Iacoboni, M., Bookheimer, S. Y., et al. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28–30.

de Gelder, B., & Hadjikhani, N. (2006). Non-conscious recognition of emotional body language. *Neuroreport*, *17*, 583–586.

Deputte, B. L. (1994). Ethological study of yawning in primates: Quantitative analysis and study of causation in two species of old world monkeys. *Ethology*, *98*, 221–245.

Goldman, A. (2005). *Simulating minds*. New York: Wiley.

Goodman, R. (2005). Infantile autism: A syndrome of multiple primary deficits? *Journal of Autism and Developmental Disorders*, *19*, 409–424.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism. *Journal of Autism and Developmental Disorders*, *39*, 693–705.

Hatfield, E., Caccioppo, J., & Rapson, R. (1994). *Emotional contagion*, New York: Cambridge University Press.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*, 253–261.

Heyes, C. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, *22*, 233–240.

Hoffman, M. L. (1978). *Empathy: Its development and prosocial implications*. Lincoln: University of Nebraska Press.

Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Development*, *67*, 1952–1969.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59,* 809–816.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27,* 145–162.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30,* 205–223.

McIntosh, D. N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion, 20,* 121–147.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary mimicry of emotional facial expressions in autism. *Developmental Science, 9,* 295–302.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198,* 75–78.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion, 7,* 447–457.

Niedenthal, P., Barsalou, L. W., Ric, F., & Krauth-Gruber, S. (2005). Embodiment in the acquisition and use of emotion knowledge. In L. F. Barrett, P. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 21–50). New York: Guilford.

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *Neuroimage, 15,* 2076–2085.

Piaget, J. (1951). *Play, dreams and imitation in childhood.* New York: Norton.

Platek, S. M., Critton, S. R., Myers, T. E., & Gallup, G. G., Jr. (2003). Contagious yawning: The role of self-awareness and mental state attribution. *Cognitive Brain Research, 17,* 223–227.

Platek, S. M., Mohamed, F. B., & Gallup, G. G., Jr. (2005). Contagious yawning and the brain. *Cognitive Brain Research, 23,* 448–452.

Provine, R. R. (1986). Yawning as a stereotyped action pattern and releasing stimulus. *Ethology, 72,* 109–122.

Provine, R. R. (1989). Faces as releasers of contagious yawning: An approach to face detection using normal human participants. *Bulletin of the Psychonomic Society, 27,* 211–214.

Provine, R. R. (1996). Contagious yawning and laughter: Significance for sensory feature detection, motor pattern generation, imitation, and the evolution of social behavior. In C. Heyes & B. Galef (Eds.), *Social learning and animals: The roots of culture* (pp. 179–208). San Diego, CA: Academic Press.

Provine, R. R. (2000). *Laughter.* New York: Penguin.

Provine, R. R. (2005). Yawning. *American Scientist, 93,* 532–539.

Reynolds, C. R. & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rogers, S. J. (2006). Studies of imitation in early infancy. In S. Rogers & J. Williams (Eds.), *Imitation and the social mind: Autism and typical development.* (pp. 3–26). New York: Guilford.

Rogers, S. J. & Williams, J. H. G. (2006). *Imitation and the social mind: Autism and typical development.* New York: Guilford.

Roid, G. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside.

Scambler, D. J., Hepburn, S. L., Rutherford, M. D., Wehner, E. A., & Rogers, S. J. (2006). Emotional responsivity in children with autism, children with other developmental disabilities, and children with typical development. *Journal of Developmental Disorders, 37,* 553–563.

Schurmann, M., Hesse, M. D., Stephan, K. E., Saarela, M., Zilles, K., Hari, R., et al. (2004). Yearning to yawn: The neural basis of contagious yawning. *Neuroimage, 24,* 1260–1264.

Senju, A., Maeda, M., Kikuchi, Y., Hasegawa, T., Tojo, Y., & Osanai, H. (2007). Absence of contagious yawning in children with autism spectrum disorder. *Biology Letters, 3,* 706–708.

Simner, M. L. (1971). Newborn's response to the cry of another infant. *Developmental Psychology, 5,* 136–150.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. (2004). Empathy for pain involves the affective but not sensory component of pain. *Science, 303,* 1157–1162.

Sparrow, S., Balla, D. A., & Cicchetti, D. V. (1984). *The Vineland Adaptive Behavior Scales–Interview Edition.* Circle Pines, MN: American Guidance Service.

Termine, N., & Izard, C. E. (1988). Infants' responses to their mother's expressions of joy and sadness. *Developmental Psychology, 24,* 223–229.

Tinbergen, N. (1951). *The study of instinct.* New York: Oxford University Press.

Troseth, G. L., & deLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development, 69,* 950–965.

Want, S. C., & Harris, Paul. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of "imitation" in children. *Developmental Science, 5,* 1–13.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 4,* 655–664.

Zahn-Waxler, C., Radke-Yarrow, M., & King, R. A. (1979). Child rearing and children's prosocial initiations toward victims of distress. *Child Development, 50,* 319–330.

Child Development, September/October 2010, Volume 81, Number 5, Pages 1632–1636

Manuscripts Accepted for Publication

An Ecological Approach to Promoting Early Adolescent Mental Health and Social Adaptation: Family Centered Intervention in Public Middle Schools, Elizabeth Stormshak, Arin M. Connell, Marie-Helene Veronneau, Michael W. Myers, Thomas J. Dishion, Kathryn Kavanagh, and Allison S. Caruthers.

Assessing the Effects of a Work-Based Antipoverty Program for Parents on Youth's Future Orientation and Employment Experiences, Rachel Kaplan, Vonnie C. McLoyd, Kelly Purtell, and Aletha Huston.

Prospective Associations Between Peer Victimization and Aggression, Jamie Ostrov.

Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning, Annie Bernier, Stephanie M. Carlson, Stéphanie Bordeleau, and Julie Carrier.

Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading, Serena Lecce, Silvia Zocchi, Adriano Pagnin, Paola Palladino, and Mele Taumoepeau.

Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups, Daphna Bassok.

Adolescents' Development of Skills for Agency in Youth Programs: Learning to Think Strategically, Reed Larson and Rachel Angus.

Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons, Olga Kochukhova and Gustaf Gredebäck.

Intergenerational Continuity in Child Maltreatment: Mediating Mechanisms and Implications for Prevention, Lisa Berlin, Karen Appleyard, and Kenneth Dodge.

Race Salience and Essentialist Thinking in Racial Stereotype Development, Kristin Pauker, Nalini Ambady, and Evan Apfelbaum.

Differing Levels of Gender Salience in Preschool Classrooms: Effects on Childrens Gender Attitudes and Intergroup Bias, Lacey J. Hilliard and Lynn S. Liben.

Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children, Geraldine Legendre, Isabelle Barriere, Louise Goyet, and Thierry Nazzi.

Two-Year-Olds Use the Generic/Nongeneric Distinction to Guide Their Inferences About Novel Kinds, Susan Graham, Samantha Nayer, and Susan Gelman.

Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children, Stuart Marcovitch, Janet J. Boseovski, Robin J. Knapp, and Michael J. Kane.

The Social Costs of Academic Success Across Ethnic Groups, Thomas E. Fuller-Rowell and Stacey N. Doan.

Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Edentity, Thomas Lyon, Elizabeth A. Ahern, Lindsay A. Malloy, and Jodi A. Quas.

Understanding Bullying and Victimization During Childhood and Adolescence: A Mixed Methods Study, Nancy Guerra, Kirk R. Williams, and Shelly Sadek.

A Developmental Perspective on Executive Function, John R. Best and Patricia Miller.

Pathways to Mathematics: Longitudinal Predictors of Performance, Jo-Anne LeFevre, Lisa Fast, Sheri-Lynn Skwarchuk, Brenda Smith-Chant, Jeffrey Bisanz, Deepthi Kamawar, and Marcie Penner-Wilger.

Development in Children's Comprehension of Linguistic Register, Laura Wagner, Maia Greene-Havas, and Rebecca Gillespie.

How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition, Clarissa Thompson and John Opfer.

Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping, Margarita Svetlova, Sara R. Nichols, and Celia A. Brownell.

A Large-Scale Evaluation of the KiVa Anti-Bullying Program, Antti Kärnä, Marinus Voeten, Elisa Poskiparta, Todd D. Little, Anne Kaljonen, and Christina Salmivalli.

Movement Planning Reflects Skill Level and Age Changes in Toddlers, Yu-ping Chen, Rachel Keen, Kerstin Rosander, and Claes von Hofsten.

Young Children Selectively Avoid Helping People With Harmful Intentions, Amrisha Vaish, Malinda Carpenter, and Michael Tomasello.

Danish Majority Children's Reasoning About Exclusion Based on Gender and Ethnicity, Signe Juhl Moller and Harriet Tenenbaum.

How Does It Look? Level 2 Perspective Taking at 36 Months of Age, Henrike Moll and Andrew Meltzoff.

Emotion Discourse, Social Cognition, and Social Skills in Children With and Without Developmental Delays, Rachel M. Fenning, Bruce L. Baker, and Jaana Juvonen.

A Monozygotic Twin Differences Study of Friends' Aggression and Children's Adjustment Problems, Frank Vitaro, Mara Brendgen, Michel Boivin, Stephane Cantin, Ginette Dionne, Richard E. Tremblay, Alain Girard, and Daniel Perusse.

Gender Differences in Adolescents' Daily Interpersonal Events and Well-Being, Lisa Flook.

Teacher–Child Relationships and the Development of Academic and Behavioral Skills During Elementary School: A Within- and Between-Child Analysis, Carolina Maldonado-Carreño and Elizabeth Votruba-Drzal.

The Effects of the Fast Track Preventive Intervention on the Development of Conduct Disorder Across Childhood, Conduct Problems Prevention Research Group.

Object Similarity Bootstraps Young Children to Action-Based Verb Extensions, Etsuko Haryu, Mutsumi Imai, and Hiroyuki Okada.

The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions, Joseph Durlak, Roger Weissberg, Allison Dymnicki, Rebecca Taylor, and Kriston Schellinger.

Science Does Not Speak for Itself: Translating Child Development Research for the Public and Its PolicyMakers, Jack Shonkoff and Susan Nall Bales.

Domain General Constraints on Statistical Learning, Erik Thiessen.

Reciprocal Relations Between Parenting and Adjustment in a Sample of Juvenile Offenders, Lela Rankin Williams and Laurence Steinberg.

A Randomized Controlled Trial of Child FIRST: A Comprehensive Home-Based Intervention Translating Research Into Early Childhood Practice, Darcy I. Lowell, Alice S. Carter, Leandra Godoy, Belinda Paulicin, and Margaret J. Briggs-Gowan.

Latino Adolescents' Experiences of Discrimination Across the First 2 Years of High School: Correlates and Influences on Educational Outcomes, Aprile Benner and Sandra Graham.

The Generic/Nongeneric Distinction Influences How Children Interpret New Information About Social Others, Andrei Cimpian and Ellen M. Markman.

Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research, Stephanie Jones, Joshua Brown, and J. Lawrence Aber.

Conditional Reasoning With False Premises Facilitates the Transition Between Familiar and Abstract Reasoning, Henry Markovits and Hugues Lortie Forgues.

When Easy Comes Hard: The Development of Adaptive Strategy Selection, Rui Mata, Bettina von Helversen, and Joerg Rieskmap.

Coherent Motion Sensitivity and Reading Development in the Transition from Prereading to Reading Stage, Bart Boets, Maaike Vandermosten, Piers Cornelissen, Jan Wouters, and Pol Ghesquière.

Same-Ethnic Friendship Preferences Among German and Turkish Preadolescents Attending Ethnically Heterogeneous Schools, Philipp Jugert, Peter Noack, and Adam Rutland.

Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect, Chiara Turati, Elisa Di Giorgio, Lara Bardi, and Francesca Simion.

Math-Gender Stereotypes in Elementary School Children, Dario Cvencek, Andrew N. Meltzoff, and Anthony G. Greenwald.

Reporter Discrepancies Among Parents, Adolescents, and Peers: Adolescent Attachment and Informant Depressive Symptoms as Explanatory Factors, Katherine B. Ehrlich, Jude Cassidy, and Matthew J. Dykas.

Urban and Rural Chinese Adolescents' Judgments and Reasoning About Personal and Group Jurisdiction, Charles Helwig, Shaogang Yang, Dingliang Tan, Chunqiong Liu, and Tiffany Shao.

Childhood Trauma and Psychiatric Disorders as Correlates of School Dropout in a National Sample of Young Adults, Michelle V. Porche, Lisa R. Fortuna, Julia Lin, and Margarita Alegria.

The Role of Peer Relationships in the Development of Early School-Age Externalizing Problems, Cristina Sturaro, Pol A. C. van Lier, Pim Cuijpers, and Hans M. Koot.

Relation Between Language Experiences in Preschool Classrooms and Children's Kindergarten and Fourth-Grade Language and Reading Abilities, David K. Dickinson and Michelle V. Porche.

Intergenerational Transmission of Adaptive Functioning: A Test of the Interactionist Model of SES and Human Development, Thomas Schofield, Monica Martin, Katherine Conger, Brent Donnellan, Tricia Neppl, and Rand Conger.

Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring, Carla Herrera, Jean Baldwin Grossman, Tina J. Kauh, and Jennifer McMaken.

Remission of Depression in Parents: Links to Healthy Functioning in Their Children, Judy Garber, Jeffrey A. Ciesla, Elizabeth McCauley, Guy Diamond, and Kelly A. Schloredt.

Double Jeopardy: Poorer Social-Emotional Outcomes for Children in the NICHD SECCYD Experiencing Home and Child Care Environments That Confer Risk, Sarah Watamura, Deborah Phillips, Taryn Morrissey, Kathleen McCartney, and Kristen Bub.

Imagining a Way Out of the Gravity Bias: Preschoolers Can Visualize the Solution to a Spatial Problem, Amy Joh, Vikram Jaswal, and Rachel Keen.

Sequential Progressions in a Theory of Mind Scale: Longitudinal Perspectives, Henry Wellman, Fuxi Fang, and Candida Peterson.

Can Maltreated Children Inhibit True and False Memories for Emotional Information? Mark Howe, Sheree Toth, and Dante Cicchetti.

Parents' Involvement in Children's Learning in the United States and China: Implications for Children's Academic and Emotional Adjustment, Cecilia Cheung and Eva Pomerantz.

Four Children and One Toy: Chinese and Canadian Children Faced With Potential Conflict Over a Limited Resource, Doran French, Xinin Chen, Janet Chung, Miao Li, Dan Li, Huichang Chen, and Li Wang.

Influence of Eye Gaze on Spoken Word Processing: An ERP Study With Infants, Eugenio Parise, Andrea Handl, Letizia Palumbo, and Angela Friederici.

Infant and Maternal Sensitivity to Interpersonal Timing, Anne Henning and Tricia Striano.

Emotion Regulation Profiles, Temperament, and Adjustment Problems in Preadolescents, Maureen Zalewski, Liliana J. Lengua, Anna C. Wilson, Anika Trancik, and Alissa Bazinet.

Developmental Pathways to Integrated Social Skills: The Roles of Parenting and Early Intervention in Raising Healthy Children, Catherine Ayoub, Claire Vallotton, and Ann Mastergeorge.

Consequences of "Minimal" Group Affiliations in Children, Yarrow Dunham, Andrew Baron, and Susan Carey.

Revisiting the Impact of Part-Time Work on Adolescent Adjustment: Distinguishing Between Selection, Kathryn C. Monahan, Joanna Lee, and Laurence Steinberg.

Parents as Moderators of the Impact of School Norms and Peer Influences on Aggression in Middle School Students, Albert Farrell, David B. Henry, Sally Mays, and Michael E. Schoeny.

The ABCs of Family Mealtimes: Observational Lessons for Promoting Healthy Outcomes for Children With Persistent Asthma, Barbara Fiese, Marcia A. Winter, and Joanna Corra Botti.

Language Development in Internationally Adopted Children: A Special Case of Early Second Language Learning, Karine Gauthier and Fred Genesee.

Third-Party Social Interaction and Word Learning From Video, Kate O'Doherty, Georgene L. Troseth, Priya Shimpi, Elizabeth Goldenberg, Nameera Akhtar, and Megan M. Saylor.

Computer-Assisted Remedial Reading Intervention for School Beginners at Risk for Reading Disability, Nina Saine, Marja-Kristiina Lerkkanen, Timo Ahonen, Asko Tolvanen, and Heikki Lyytinen.

Desire for Increased Autonomy and Adolescents' Perceptions of Peer Autonomy: "Everyone Else Can; Why Can't I?" Christopher Daddis.

Almost Thinking Counterfactually: Children's Understanding of Close Counterfactuals, Sarah Beck and Carlie Guthrie.

CSRP's Impact on Low-Income Preschoolers' Pre-Academic Skills: Self-Regulation as a Mediating Mechanism, C. Cybele Raver, Stephanie M. Jones, Christine Li-Grining, Fuhua Zhai, Kristen Bub, and Emily Pressler.

Does God Make It Real? Children's Belief in Religious Stories From the Judeo-Christian Tradition, Victoria Cox Vaden and Jacqueline D. Woolley.

Changes in Early Adolescents' Sense of Responsibility to Their Parents in the United States and China: Implications for Their Academic Functioning, Eva Pomerantz, Lili Qin, Qian Wang, and Huichang Chen.

Two- to Eight-Month-Old Infants' Perception of Dynamic Auditory–Visual Spatial Colocation, Gavin Bremner, Alan Slater, Scott Johnson, Uschi Mason, Jo Spring, and Maggie Bremner.

Longitudinal Associations Between Marital Instability and Child Sleep Problems Across Infancy and Toddlerhood in Adoptive Families, Anne Mannering, Gordon Harold, Leslie Leve, Katherine Shelton, Daniel Shaw, Rand Conger, Jenae Neiderhiser, Laura Scaramella, and David Reiss.

Promoting Effective Parenting Practices and Preventing Child Behavior Problems in School, Laurie Brotman, Esther Calzada, Keng-Yen Huang, Sharon Kingston, Spring Dawson-McClure, Dimitra Kamboukos, Amanda Rosenfelt, Ami Schwab, and Eva Petkova.

Age 26 Cost–Benefit Analysis of the Child-Parent Center Early Education Program, Arthur Reynolds, Judy Temple, Barry White, Suh-Ruu Ou, and Dylan Robertson.

Infantile Amnesia Across the Years: A 2-Year Follow-Up of Children's Earliest Memories, Carole Peterson, Kelly L. Warren, and Megan M. Short.

Hemoglobin, Growth, and Attention of Infants in Southern Ethiopia, Nicki Aubuchon-Endsley, Stephanie L. Grant, Getenesh Berhanu, David G. Thomas, Sarah Schrader, Devon Eldridge, Tay S. Kennedy, and Michael Hambidge.

Transition From Crawling to Walking and Infants' Actions With Objects and People, Lana Karasik, Catherine Tamis-LeMonda, and Karen Adolph.

The Genetic and Environmental Covariation Among Psychopathic Personality Traits and Reactive and Proactive Aggression in Childhood, Serena Bezdjian, Catherine Tuvblad, Adrian Raine, and Laura A. Baker.

Structure Mapping and Relational Language Support Children's Learning of Relational Categories, Dedre Gentner, Florencia K. Anggoro, and Raquel S. Klibanoff.

Narrative Elaboration and Participation: Two Dimensions of Maternal Elicitation Style, Gigliana Melzi, Adina R. Schick, and Joy L. Kennedy.

Does Felt Gender Compatibility Mediate Influences of Self-Perceived Gender Nonconformity on Early Adolescents' Psychosocial Well-Being? Meenakshi Menon.

Theory of Mind and Social Interest in Zero Acquaintance Play Situations, Chris Moore, Sandra Bosacki, and Shannon Macgillivray.

Media as Social Partners: The Social Nature of Young Children's Learning From Screen Media, Rebekah Richert, Michael B. Robb, and Erin I. Smith.

Protecting Children From the Consequences of Divorce: A Longitudinal Study of the Effects of Parenting on Children's Coping Processes, Clorinda Velez, Sharlene Wolchik, Jenn-Yun Tein, and Irwin Sandler.

Preschoolers' Quarantining of Fantasy Stories, Rebekah Richert and Erin I. Smith.

Effect of Foster Care on Young Children's Language Learning, Jennifer Windsor, Joann Benigno, Christine Wing, Patrick Carroll, Sebastian Koga, Charles Nelson, Nathan Fox, and Charles Zeanah.

Why Don't Smart Teens Have Sex? A Behavioral Genetic Approach, Kathryn Paige Harden and Jane Mendle.

Preschoolers' Use of Morphosyntactic Cues to Identify Generic Sentences: Indefinite Singular Noun Phrases, Tense, and Aspect, Andrei Cimpian, Trent J. Meltzer, and Ellen M. Markman.

Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability, Michele Mazzocco, Lisa Feigenson, and Justin Halberda.

Imitation in Infancy: Rational or Motor Resonance? Markus Paulus, Sabine Hunnius, Marlies Vissers, and Harold Bekkering.

Effects of Nonmaternal Care in the First 3 Years on Children's Academic Skills and Behavioral Functioning in Childhood and Early Adolescence, Sara Jaffee, Carol Van Hulle, and Joseph Rodgers.

Shyness-Sensitivity and Unsociability in Rural Chinese Children: Relations With Social, School, and Psychological Adjustment, Xinyin Chen and Li Wang.

Age Differences in Strategic Planning as Indexed by the Tower of London, Dustin Albert and Laurence Steinberg.

Young Children's Learning and Transfer of Biological Information From Picture Books to Real Animals, Patricia Ganea, Lili Ma, and Judy DeLoache.

Characterizing Children's Expectations About Expertise and Incompetence: Halo or Pitchfork Effects? Melissa Koenig and Vikram K. Jaswal.

Wherein Lies Children's Intergroup Bias? Egocentrism, Social Understanding, and Social Projection, Dominic Abrams.

Can Children With Autism Spectrum Disorders "Hear" a Speaking Face? Julia Irwin, Lawrence Brancazio, Lauren Tornatore, and D. H. Whalen.

The Home Learning Environment Across the First 5 Years: Associations With Children's Vocabulary and Emergent Literacy Skills at Pre-Kindergarten, Eileen T. Rodriguez and Catherine S. Tamis-LeMonda.

A Longitudinal Examination of African American Adolescents' Attributions About Achievement Outcomes, Akilah Swinton, Beth Kurtz-Costes, Stephanie Rowley, and Ndidi Okeke.

Ethnic Stigma, Academic Anxiety, and Intrinsic Motivation in Middle Childhood, Cari Gillen-O'Neel, Diane N. Ruble, and Andrew J. Fuligni.

Examining the Black–White Achievement Gap Among Low-Income Children: Using the NICHD Study of Early Child Care and Youth Development, Margaret Burchinal, Kathleen McCartney, Laurence Steinberg, Robert Crosnoe, Sarah L. Friedman, Vonnie McLoyd, and Robert Pianta.

A Longitudinal Study of Religious Identity and Participation During Adolescence, Anna Lopez, Virginia Huynh, and Andrew Fuligni.

Benefits of Practicing 4 = 2 + 2: Nontraditional Problem Formats Facilitate Children's Understanding of Mathematical Equivalence, Nicole M. McNeil, Emily R. Fyfe, Lori A. Petersen, April E. Dunwiddie, and Heather Brletic-Shipley.

Reciprocal Associations Between Family and Peer Conflict in Adolescents' Daily Lives, Grace Chung, Lisa Flook, and Andrew Fuligni.

The Development of Distrust, Kimberly Vanderbilt, David Liu, and Gail Heyman.

Developing Relationships, Being Cool, and Not Looking Like a Loser: Social Goal Orientation Predicts Children's Responses to Peer Aggression, Karen D. Rudolph, Jamie L. Abaied, Megan Flynn, Niwako Sugimura, and Anna Monica Agoston.

Nonlinear Growth Curves in Developmental Research, Kevin Grimm, Nilam Ram, and Fumiaki Hamagami.

Developmental Continuity in Theory of Mind: Speed and Accuracy of Belief-Desire Reasoning in Children and Adults, Ian Apperly, Frances Warren, Benjamin Andrews, Jay Grant, and Sophie Todd.

Children's Judgments of Emotion From Conflicting Cues in Speech: Why 6-Year-Olds Are So Inflexible, Matthew Waxer and J Morton.

Social-Emotional Effects of Early Childhood Education Programs in Tulsa, William Gormley Jr., Deborah A. Phillips, Katie Newmark, Kate Perper, and Shirley Adelstein.

Society for Research in Child Development
IS Developmental Science

SRCD Exchange
exchange.srcd.org

SRCD's Social Media Site
Connect with your colleagues in the field!

Top 5 reasons why YOU need to be a SRCD member...

1. Play an Active Role & Broaden Your Horizons in the Society & the Field
2. Link, Share, & Connect with Policy Matters
3. Be "In the Know"
4. Opportunities to Apply for Grants & Awards
5. Many Member Extras

Join SRCD Today!
www.srcd.org

Questions regarding SRCD Membership contact: membership@srcd.org

Education • Linguistics • Neuroscience • Philosophy • Psychology • Anthropology • Artificial Intelligence

Volume 33, Number 1, January/February, 2009 ISSN 0364-0213

COGNITIVE SCIENCE

A Multidisciplinary Journal

Journal of the Cognitive Science Society

Executive Editor: Arthur B. Markman

Associate Editors: John R. Anderson • Michael D. Byrne
Susan Goldin-Meadow • Lawrence A. Hirschfeld • Michael D. Lee
James S. Magnuson • Danielle S. McNamara • Johanna D. Moore
Nancy J. Nersessian • Mark A. Pitt • Rajesh Rao • Terry Regier

WILEY-BLACKWELL

COGNITIVE
SCIENCE

Edited by
ARTHUR B. MARKMAN

Published on behalf of the Cognitive Science Society

Cognitive Science publishes articles in all areas of cognitive science, covering such topics as knowledge representation, inference, memory processes, learning, problem solving, planning, perception, natural language understanding, connectionism, brain theory, motor control, intentional systems, and other areas of interdisciplinary concern. Highest priority is given to research reports that are specifically written for a multidisciplinary audience. The audience is primarily researchers in cognitive science and its associated fields, including anthropologists, education researchers, psychologists, philosophers, linguists, computer scientists, neuroscientists, and roboticists.

WILEY-
BLACKWELL

For more information and to subscribe online visit
wileyonlinelibrary.com/journal/cogs

JOURNAL OF

Personality

Volume 77
Number 3
June
2009

Geneviève A. Mageau, Robert J. Vallerand, Julie Charest, Sarah-Jeanne Salvy, Nathalie Lacaille, Thérèse Bouffard, and Richard Koestner
On the Development of Harmonious and Obsessive Passion: The Role of Autonomy Support, Activity Specialization, and Identification With the Activity 601

Ruthellen Josselson
The Present of the Past: Dialogues With Memory Over Time 647

Keiko A. Taga, Howard S. Friedman, and Leslie R. Martin
Early Personality Traits as Predictors of Mortality Risk Following Conjugal Bereavement 669

Todd B. Kashdan, Anjali Mishra, William E. Breen, and Jeffrey J. Froh
Gender Differences in Gratitude: Examining Appraisals, Narratives, the Willingness to Express Emotions, and Changes in Psychological Needs 691

Joshua Weller and David Watson
Friend or Foe? Differential Use of the Self-Based Heuristic as a Function of Relationship Satisfaction 731

Joshua D. Miller, W. Keith Campbell, Diana L. Young, Chad E. Lakey, Dennis E. Reidy, Amos Zeichner, and Adam S. Goodie
Examining the Relations Among Narcissism, Impulsivity, and Self-Defeating Behaviors 761

Susan J. Wenze, Kathleen C. Gunthert, Nicholas R. Forand, and Jean-Philippe Laurenceau
The Influence of Dysphoria on Reactivity to Naturalistic Fluctuations in Anger 795

Peter Borkenau, Katrin Zaltauskas, and Daniel Leising
More May Be Better but There May Be Too Much: Optimal Trait Level and Self-Enhancement Bias 825

Thomas W. Britt, Matthew R. Millard, Preetha T. Sundareswaran, and DeWayne Moore
Personality Variables Predict Strength-Related Attitude Dimensions Across Objects 859

Eamonn Ferguson, Peter A. Bibby, Sara Rosamond, Claire O'Grady, Alison Parcell, Christopher Amos, Christine McCutcheon, and Ronan O'Carroll
Alexithymia, Cumulative Feedback, and Differential Response Patterns on the Iowa Gambling Task 883

Wiley-Blackwell, Boston, MA and Oxford, UK

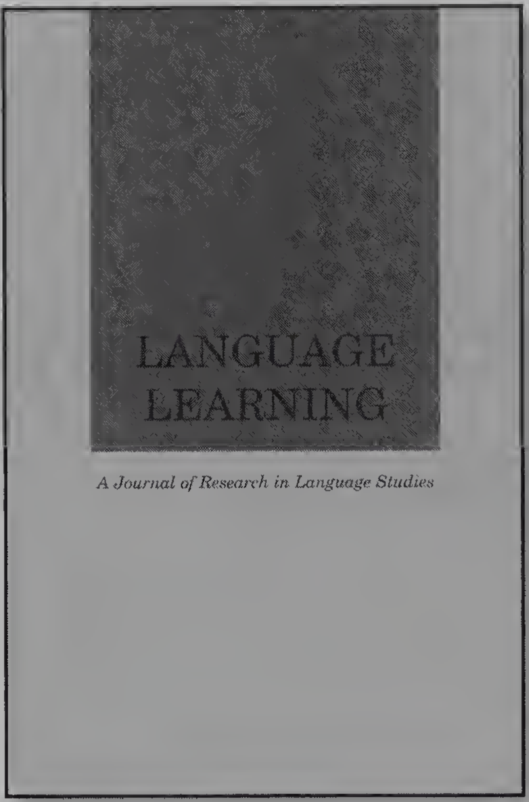
JOURNAL OF
Personality

Edited by
HOWARD TENNEN

Journal of Personality publishes scientific investigations in the field of personality. It focuses particularly on personality and behavior dynamics, personality development, and individual differences in the cognitive, affective, and interpersonal domains. The journal reflects and stimulates interest in the growth of new theoretical and methodological approaches in personality psychology.

WILEY-
BLACKWELL

For more information and to subscribe online visit
wileyonlinelibrary.com/journal/jopy



LANGUAGE
LEARNING

A Journal of Research in Language Studies

LANGUAGE LEARNING

A Journal of Research in Language Studies

Edited by
NICK C. ELLIS *(General Editor),*
LOURDES ORTEGA *(Journal Editor),*
MARIANNE GULLBERG *(Associate Journal Editor),*
SCOTT JARVIS *(Associate Journal Editor),*
ALISTER H. CUMMING *(Executive Director and Best of Language Learning Series Editor),*
MARY J. SCHLEPPEGRELL *(Monograph Series Editor),*
JOHN H. SCHUMANN *(LL-MPI Cognitive Neuroscience Series Editor)*

Published on behalf of the Language Learning Research Club at the University of Michigan

Language Learning is a scientific journal dedicated to the understanding of language learning broadly defined. It publishes research articles that systematically apply methods of inquiry from disciplines including psychology, linguistics, cognitive science, educational inquiry, neuroscience, ethnography, sociolinguistics, sociology, and semiotics. It is concerned with fundamental theoretical issues in language learning such as child, second, and foreign language acquisition, language education, bilingualism, literacy, language representation in mind and brain, culture, cognition, pragmatics, and intergroup relations. A subscription includes an annual supplement – a volume from the ***Best of Language Learning Series*** or the ***Language Learning Monograph Series*** – as well as a biennial monograph – the ***Language Learning-Max Planck Institute Cognitive Neurosciences Series.***

WILEY-
BLACKWELL

For more information and to subscribe online visit
wileyonlinelibrary.com/journal/lang

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

FIRST-YEAR MATERNAL EMPLOYMENT AND CHILD DEVELOPMENT IN THE FIRST 7 YEARS
- *Jeanne Brooks-Gunn, Wen-Jui Han, and Jane Waldfogel* (Serial No. 296, 2010).

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION
- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial No. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY
- *Mona El-Sheikh et al.* (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN
- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING
- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE
- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS
- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE
- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT
- *Michelle M. Chouinard* (Serial No. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS
- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM
- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997
- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD
- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS
- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT
- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES
- *Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES
- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR
- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

WILEY-
BLACKWELL

NOVEMBER/DECEMBER 2010 • VOLUME 81, NUMBER 6

CHILD DEVELOPMENT

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

SRCD
Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit wileyonlinelibrary.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives.* Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at wileyonlinelibrary.com/journal/cdev. This journal is available online at Wiley Online Library. Visit wileyonlinelibrary.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to www.wileycustomerhelp.com/ask or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.

Visit our Online Customer Get-Help available in 7 languages at www.wileycustomerhelp.com

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Child Development © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor
Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor
Detra N. Davis

Editorial Assistant
Adam Martin

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Ramesh Bhatt
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Cara Cashon
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Catherine Costigan
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Heidi Gazelle
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Wendy Hoglund
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Kristin H. Lagattuta
Richard Lee
Charlie Lewis
Jeffrey Lidz
Ramaswami Mahalingam
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Marelene Moretti
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Jenae Neiderhiser
Drew Nesdale
Nora S. Newcombe
Candice L. Odgers
David Oppenheim
Anna Papafragou
Kristin Pauker
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Cybele C. Raver
Robert Reeve
Greg Reynolds
John J. Rieser
Kristen Ritchey
Deborah Rivas-Drake
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Eleanor Seaton
Rushen Shi
Judi G. Smetana
Alan L. Smith
Catherine Snow
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, November/December 2010, Volume 81, Number 6, Pages 1637–1924

Contents

IN THIS ISSUE 1637

REVIEW

A Developmental Perspective on Executive Function
John R. Best and Patricia H. Miller 1641

EMPIRICAL REPORTS

Young Children Selectively Avoid Helping People With Harmful Intentions
Amrisha Vaish, Malinda Carpenter, and Michael Tomasello 1661

Prospective Associations Between Peer Victimization and Aggression
Jamie M. Ostrov 1670

Development in Children's Comprehension of Linguistic Register
Laura Wagner, Maia Greene-Havas, and Rebecca Gillespie 1678

Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children
Stuart Marcovitch, Janet J. Boseovski, Robin J. Knapp, and Michael J. Kane 1687

EMPIRICAL ARTICLES

The Social Costs of Academic Success Across Ethnic Groups
Thomas E. Fuller-Rowell and Stacey N. Doan 1696

Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Identity
Thomas D. Lyon, Elizabeth C. Ahern, Lindsay C. Malloy, and Jodi A. Quas 1714

Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons
Olga Kochukhova and Gustaf Gredebäck 1729

Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning
Annie Bernier, Stephanie M. Carlson, Stéphanie Bordeleau, and Julie Carrier 1739

Pathways to Mathematics: Longitudinal Predictors of Performance
Jo-Anne LeFevre, Lisa Fast, Sheri-Lynn Skwarchuk, Brenda L. Smith-Chant, Jeffrey Bisanz, Deepthi Kamawar, and Marcie Penner-Wilger 1753

How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition
Clarissa A. Thompson and John E. Opfer 1768

Differing Levels of Gender Salience in Preschool Classrooms: Effects on Children's Gender Attitudes and Intergroup Bias
Lacey J. Hilliard and Lynn S. Liben 1787

Race Salience and Essentialist Thinking in Racial Stereotype Development
Kristin Pauker, Nalini Ambady, and Evan P. Apfelbaum 1799

Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping
Margarita Svetlova, Sara R. Nichols, and Celia A. Brownell 1814

Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups
Daphna Bassok 1828

Movement Planning Reflects Skill Level and Age Changes in Toddlers
Yu-ping Chen, Rachel Keen, Kerstin Rosander, and Claes von Hofsten 1846

Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children
Géraldine Legendre, Isabelle Barrière, Louise Goyet, and Thierry Nazzi 1859

Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading
Serena Lecce, Silvia Zocchi, Adriano Pagnin, Paola Palladino, and Mele Taumoepeau 1876

Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect
Chiara Turati, Elisa Di Giorgio, Lara Bardi, and Francesca Simion 1894

OTHER

Manuscripts Accepted 1906

Editorial Acknowledgments 1911

Index to Volume 1917

Child Development, November/December 2010, Volume 81, Number 6, Pages 1637–1640

In This Issue

Executive Functioning

Best and Miller (p. 1641) reviewed the research on children's executive functions—cognitive skills that enable children and adolescents to control their thoughts and behavior and to act in a goal-directed manner when they are in a new situation. Their examination goes beyond prior reviews, which focused on preschoolers, to include elementary schoolers and teens. Executive function is a critical part of children's development because it is tied to their ability to regulate impulsive behavior and to perform academically. The authors conclude that a developmental perspective is critical for understanding changes in executive functions and their relations over time, as these abilities emerge early in life—in infancy—and continue to mature through adolescence.

Executive functioning—including such skills as impulse control, working memory, and mental flexibility—allows people to consciously plan and monitor their behavior so they can reach desired goals. Why are some children better at these skills than others? In their longitudinal study of 60 Canadian children at ages 1, 1½, and 2, **Bernier, Carlson, Bordeleau, and Carrier** (p. 1739) find that nighttime sleep plays a role. Children who got most of their sleep at night (as opposed to during the day) at ages 1 and 1½ did better in these skill areas than children who did not sleep as much at night. The findings suggest that infant sleep has an effect on developing brain structures in the first 2 years of life, thereby setting in motion a cascade of neural effects that may have implications for later executive skills.

Research has shown that adults' ability to follow through on instructions (called goal maintenance) is related to working memory capacity—the maximum number of items you can keep active in your consciousness. **Marcovitch, Boseovski, Knapp, and Kane** (p. 1687) ask: Is successful goal maintenance in sixty-five 4- and 6-year-olds related to their working memory capacity, or do children at these ages simply lack the ability to keep rules in mind regardless of their working memory? They conclude that children with higher working memory scores were more likely to maintain goals, suggesting that higher order cognitive ability is built from lower order components, even at this young age. The study demystifies parts of executive function and provides insights into the skills needed to engage in the conscious control of behavior, which is a key aspect of success in life.

Helping Hands

Children begin to behave prosocially—that is, they factor others into their actions—before they turn 2. In a study of 65 mostly White 18- and 30-month-olds, **Svetlova, Nichols, and Brownell** (p. 1814) sought to learn why. The researchers find that emphatic and altruistic ability may not be in place in 1-year-olds, even though some of their behaviors may indicate otherwise. They also find that the same behavior can have different meanings at different periods of children's development, with the meaning of helping, in particular, changing over the first few years to include helping that is instrumental (based on action), empathic (based on emotion), and altruistic (requiring a child to give up something that belongs to him or her). Early prosocial behavior, it appears, develops over the 2nd year as children slowly move from instrumental helping to the more complex ability to understand and act on others' specific needs. The study can help adults understand how to encourage young children to behave prosocially.

Vaish, Carpenter, and Tomasello (p. 1661) asked: When young children act prosocially, do they help everyone equally, or are they less likely to help those with harmful intentions? In two studies, of almost 100 German 3-year-olds, they find: Children as young as 3 helped an adult less not only when they saw that the person harmed another person but also when they saw that the person *meant* to harm another person without causing actual harm. And they helped an adult when that person accidentally caused harm. In finding that children are quite sophisticated and discriminating helpers, the study shows that youngsters are sensitive not only to others' moral behaviors but also to the intentions behind those behaviors.

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0001

Aggression and Aggressors

Do young children who are physically victimized—hit, kicked, or pushed—by their classmates learn from these interactions to become more physically aggressive over time? Do preschoolers who are socially excluded by their peers model these behaviors and engage in the same type of aggression later? **Ostrov** (p. 1670) took a longitudinal look at more than 100 preschoolers to conclude that young children who are the victims of aggression become aggressors over the course of a school year, with physical victimization leading to physical aggression and relational victimization leading to relational aggression. By illustrating that children may retaliate in the same way based on the harassment they have been subjected to by peers, the findings can inform intervention programs.

Too Smart to Fit in?

Achievement and social acceptance are both important challenges during adolescence. To what extent do the two clash, with students being ostracized for being smart? **Fuller-Rowell and Doan** (p. 1696), in a longitudinal study of a nationally representative sample of almost fourteen thousand 7th through 12th graders over a year, show that ethnicity and school context are both important factors to consider when addressing this question. African American and Native American teens experienced greater social costs with achievement than Whites or other ethnic groups, even after accounting for various background characteristics. The social costs experienced by African Americans were greatest in higher achieving competitive schools, but only when there were not a substantial number of same-race peers at the schools. Similar effects were found for students of Mexican descent. The study, the first to quantitatively show differential social costs with academic success across ethnic groups, points to the significance of issues of race and inequality in schools and can inform programs to address these issues.

Full Disclosure

In one of the first studies to address how children think about telling others about adults' wrongdoing, **Lyon, Ahern, Malloy, and Quas** (p. 1714) focused on maltreated children (children who had suffered from neglect and/or physical or sexual abuse and had been removed from their parents' custody), contrasting their attitudes with a comparable group of children with no history of maltreatment. They asked almost 300 children aged 4–9 to discuss stories that involved adult wrongdoing. Even the youngest (4- and 5-year-olds) distinguished between wrongdoers who were parents and those who were strangers, more often telling when the wrongdoer was a stranger. Older children (8- and 9-year-olds) were more likely to tell one parent about the other parent's wrongdoing, keeping the information in the family. The maltreated children acted somewhat differently, with younger children disclosing less; older maltreated children disclosed more, showing less protectiveness toward parents who had done wrong than children who had not been maltreated—perhaps because of their growing distrust of their parents. The study has implications for legal proceedings involving children.

Building Blocks of Math Knowledge

Can we tell how well children will learn mathematics before they even start first grade? Does early math knowledge predict later math performance? In a longitudinal study, **LeFevre et al.** (p. 1753) tested the math abilities of more than 180 Canadian children aged 4 through 6, then tested them again 2 years later. The researchers find that three kinds of cognitive skills—quantitative, linguistic, and attentional—were related to children's early number knowledge and basic mathematical skills, and these early skills predicted how children did on tasks that assessed conventional math knowledge 2 years later. The results provide a framework for understanding the early links between children's cognitive skills and the development of mathematics knowledge. As such, they can help teachers better understand how children learn math.

According to the National Mathematics Advisory Panel, fostering children's number sense should be made a high priority. The problem is that numbers are infinite, but improvements in number sense typically occur gradually for larger numbers. To address this problem, **Thompson and Opfer** (p. 1768) investigated whether the piecemeal nature of typical number sense development could be exploited to induce broad and rapid changes in number sense. They find that with progressive alignment—when a small numeric range (such as 0–100) for which children had already produced accurate estimates was aligned with large numeric ranges (such as 0–1,000)—children drew analogies between the small and large ranges, and this led them to note deeper commonalities (such as the underlying structure of the decimal system) and produce more accurate estimates for numbers as high as 100,000. This suggests that when instruction is tailored to support analogies, children's existing knowledge about numbers can serve as a foundation for more advanced numerical understanding.

Effects of Preschool by Race

Recent research suggests that the effects of preschool on young children's cognitive development vary by race, with Black and Hispanic children showing greater benefits than their White peers. **Bassok** (p. 1828) conducted a study based on new information on about 7,400 children who were part of the Early Childhood Longitudinal Study, a nationally representative data set that tracks children from birth to kindergarten. She finds that on average, poor 4-year-olds benefit substantially from preschool participation and they do so at about the same rate, regardless of their racial and ethnic background. In contrast, among nonpoor children, the effects of preschool participation differ across groups, with Black children experiencing the biggest boosts. As the number of 4-year-olds in state preschool programs rises, this study can inform the development of programs to help narrow achievement gaps between races.

Hand and Mind

Even 6-month-olds can anticipate goals of manual actions with which they are familiar. That's what **Kochukhova and Gredebäck** (p. 1729) find in their study of infants' ability to anticipate the goal of such actions as feeding and combing hair. Using eye-tracking technology, the researchers sought to determine if infants looked at the goal of the action before it was completed. Based on their finding that babies are better able to anticipate actions involved in feeding (an activity with which they are familiar) than the actions involved in combing hair (an activity with which they are unfamiliar), the researchers conclude that the babies can visually anticipate activities with which they have experience. Furthermore, they find that this ability improves with age—10-month-olds were faster than 6-month-olds. The study, which involved more than 100 babies, sheds light on how infants perceive others' actions.

When children are figuring out how to use motor skills—learning to reach, grasp, and walk—they are also sequencing several motor acts to achieve complex goals. But they're also advancing cognitively, based on a longitudinal study in which children performed a precise task (tower building) and an imprecise task (putting blocks in a large container) by **Chen, Keen, Rosander, and von Hofsten** (p. 1846). Toddlers (aged 18–21 months) do not just carry out actions in isolation, they conclude, but rather actively plan their actions in advance, evaluating the precision demands of the task and adjusting their movement patterns accordingly. A year later, the authors find that the children had matured in their use of arm and hand movements, and individual differences in the speed of some of these movements remained the same. By drawing attention to the interplay between motor skills and cognitive ability, this study offers insights into toddlers' thought processes as they plan movements for tasks with different precision requirements.

Preschoolers' Concepts of Gender and Race

In many preschool classrooms, gender is very noticeable—think of the greeting, "Good morning, boys and girls" or the instruction, "Girls line up on the left, boys on the right." Do such environments promote children's stereotyped views of what activities and jobs are appropriate for boys and girls? Do they affect children's interest in playing with children of the other sex? **Hilliard and Liben** (p. 1787) evaluated more than fifty 3- to 5-year-olds over 2 weeks; in one set of classrooms, teachers continued to avoid making divisions by sex. In the other, teachers used gendered language and divisions (like asking boys and girls to post their work on separate bulletin boards) but still avoided statements comparing boys and girls or fostering competition between them. Children in the classrooms in which teachers avoided characterizations by sex showed no change in responses or behaviors, but children in the other classrooms had more stereotyped attitudes, said they were less interested in playing with children of the other sex, and were observed to play less with these children. By highlighting the powerful effect of classroom environments on children's gender-related beliefs and behaviors, the findings have implications for how teachers interact with children.

Knowledge of racial stereotypes can have important consequences for children: Stereotypes affect children's treatment of others and shape expectations about their own capacity for achievement. In a study of almost 90 mostly European American children aged 3–10, **Pauker, Ambady, and Apfelbaum** (p. 1799) find that children begin applying racial stereotypes at around age 6, particularly those involving individuals of a race different than their own. In addition, two key factors predict whether they use stereotypes—their ability to group others by race and their tendency to think about race as a permanent quality. By teaching children that even though a person's skin color usually cannot be changed, race does not convey stable information about a person's traits, they might be less likely to stereotype others and less susceptible to stereotypes themselves.

Language Learning and Learning About Language

Many languages encode a difference between singular and plural reference. English, for example, encodes singular nouns in one way (a boy) and plurals in another (some boys); verbs also reflect this difference. Little is known about how and when young learners understand this linguistic feature. **Legendre, Barrière, Goyet, and Nazzi** (p. 1859) carried out two experiments with about 75 French children (aged 2 and 2½) learning French. They find that children as young as 2½ (much earlier than previously thought) understand the difference between singular versus plural verb forms in the absence of any other verbal cue. The study helps define the normal course of learning a native language, which can help us understand and identify possible language-related deficits that might not otherwise be detected.

When you're involved in a conversation, you use different linguistic styles depending on whom you're talking to, with the same person speaking differently to a teacher than to a baby, for example. In a study of more than 50 mostly White 3- to 5-year-olds, **Wagner, Greene-Havas, and Gillespie** (p. 1678) sought to determine when children understand that the way people speak provides information about the person they're speaking to. Findings: Around age 4, children can listen to how someone speaks to figure out whom they are speaking to, and this ability continues to improve with age. This suggests that preschoolers are quite sophisticated in their understanding of language and social contexts: Even before they have started formal schooling, they have already learned how others signal their social relationships through subtle shifts in their linguistic style.

Reading Minds

Part of what makes us human is our ability to understand that people's behavior is motivated by what goes on in their minds. This understanding is referred to as mental state knowledge, or the awareness that mental states such as memories, beliefs, desires, and intentions govern people's behavior. **Lecce, Zocchi, Pagnin, Palladino, and Taumoepeau** (p. 1876) wanted to know how young children's knowledge about mental states relates to their understanding of how the mind works when they are older. In their study of more than one hundred fifty 7- to 10-year-olds in Italy, they examined the relation between children's knowledge of mental states and their later understanding of the reading process. They find a genuine link between children's knowledge of mental states (also known as theory of mind) and their metaknowledge about reading (their ability to understand the process of understanding text) that is causal (going from earlier mental state knowledge to later metaknowledge about reading). And they suggest that this relation is specific to children's knowledge about beliefs rather than emotion. The study has implications for educators.

Face Processing

Research on infants' reactions to faces addresses a controversial question: Is sensitivity to faces a special ability in humans? If so, how does it develop? Turati, Di Giorgio, Bardi, and Simion (p. 1894) used an innovative technique to explore whether sensitivity to faces in newborns and 3-month-olds is determined by processing the whole face (holistic processing) or individual components of the face (feature processing). Like adults, 3-month-olds recognized a face relying mainly on holistic information. Newborns, however, did not show evidence of holistic processing, which is consistent with the notion that face processing develops over time rather than exists fully formed at birth.

Anne Bridgman

Child Development, November/December 2010, Volume 81, Number 6, Pages 1641–1660

A Developmental Perspective on Executive Function

John R. Best and Patricia H. Miller
University of Georgia

This review article examines theoretical and methodological issues in the construction of a developmental perspective on executive function (EF) in childhood and adolescence. Unlike most reviews of EF, which focus on preschoolers, this review focuses on studies that include large age ranges. It outlines the development of the foundational components of EF—inhibition, working memory, and shifting. Cognitive and neurophysiological assessments show that although EF emerges during the first few years of life, it continues to strengthen significantly throughout childhood and adolescence. The components vary somewhat in their developmental trajectories. The article relates the findings to long-standing issues of development (e.g., developmental sequences, trajectories, and processes) and suggests research needed for constructing a developmental framework encompassing early childhood through adolescence.

Broadly defined, executive functions (EFs) encompass those cognitive processes that underlie goal-directed behavior and are orchestrated by activity within the prefrontal cortex (PFC; e.g., Olson & Luciana, 2008; Shimamura, 2000). Children's EF has been of great interest to developmental psychologists in recent years. However, this research has three limitations that pose difficulties for constructing a truly developmental account of EF. First, most research on the development of EF has examined narrow age ranges, for example, 2–5 (Isquith, Gioia, & Espy, 2004). Second and relatedly, most research has focused on preschoolers (e.g., Carlson, 2005; Garon, Bryson, & Smith, 2008), perhaps because rapid improvements occur during the preschool and early school years on EF tasks (e.g., Carlson & Moses, 2001; Zelazo, Muller, Frye, & Marcovitch, 2003). However, performance on other, more complex tasks does not mature until adolescence or even early adulthood (e.g., Anderson, 2002; Conklin, Luciana, Hooper, & Yarger, 2007; Davidson, Amso, Anderson, & Diamond, 2006; Luciana, Conklin, Hooper, & Yarger, 2005; Romine & Reynolds, 2005). Moreover, the rudiments of EF emerge before early childhood, likely within the 1st year of life (e.g., Diamond, 1990a, 1990b). Third, we have little information about the processes by which children move from one level to another, especially processes operating after age 5.

Consequently, despite the large literature on EF in children, we have no truly developmental account of EF across childhood and adolescence. The purpose of this article is to begin to construct such an account, which distinguishes it from previously published reviews of EF (but see Best, Miller, & Jones, 2009). We focus on the few studies that include a large age range in an attempt to outline the broad developmental trajectories of EF and look at the development of EF within the framework of developmental theoretical issues.

After a brief presentation of theoretical and methodological challenges to a developmental account of EF, the main part of the review examines changes in three components of EF across multiple ages. Then we address developmental trajectories, sequences of the components, and mechanisms of development, and suggest future research to examine basic issues of development.

Theoretical and Methodological Challenges

Beyond the limitations of narrow age ranges and few studies examining developmental sequences and mechanisms, it is very difficult for other reasons to extract a general trajectory of EF development

---

Support for this work was provided in part by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (Grant RO1 DK70922-01). We thank Lara Jones, Katherine Davis, Philip Tomporowski, and Jack Naglieri for reading and offering comments on an earlier draft.

Correspondence concerning this article should be addressed to Patricia H. Miller, Department of Psychology, University of Georgia, Athens, GA 30602. Electronic mail may be sent to phmiller@uga.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0002

from the literature. A main challenge is the lack of agreement concerning whether EF is a unitary construct or a set of independent components (e.g., Barkley, Edwards, Laneri, Fletcher, & Metevia, 2001; Brocki & Bohlin, 2004; Isquith et al., 2004; Miyake et al., 2000). One prominent theoretical framework integrates these opposing perspectives by suggesting that the EF construct consists of interrelated, but distinct, components—described as the ''unity and diversity of EF'' by Miyake et al. (2000). In their seminal study with young adults, Miyake et al. used confirmatory factor analysis (CFA) to test this framework. The CFA extracted three correlated latent variables from several commonly used EF tasks. These latent variables represented three EF components—inhibition, working memory (WM), and shifting—that contributed differentially to performance on complex EF tasks (Miyake et al., 2000; although Miyake recently—2009—has questioned whether inhibition can be considered a distinct component). Thus, although bound by some common underlying processes, in young adults EFs are distinguishable and are employed differentially based on the task at hand.

Some research with children has investigated the EF construct and has found at least partial support for an integrative framework. Hughes (1998) sought to expand a previous finding that EF consists of dissociable components in older children (Welsh, Pennington, & Groisser, 1991). She extracted three distinct factors—attentional flexibility, inhibitory control, and WM—from preschoolers' performance on several EF tasks, suggesting that EF components are differentiated even at a young age. Both Hughes and Welsh et al. (1991) emphasize the independence of these factors, leaving little discussion of whether these factors may be interrelated. Senn, Espy, and Kaufmann (2004), also with preschoolers, used path analysis, which forms each latent variable by drawing on only one task rather than multiple tasks (which makes it more susceptible to extraneous influences such as test order and task reliability that may affect relations among the measures). Although performance on the WM and inhibition tasks was correlated and predicted complex task performance, shifting performance was unrelated to the other measures. This provided evidence that the EF components are dissociable in early childhood but also that those components are interrelated to some degree.

CFA with older children seems to provide stronger support for Miyake's ''unity and diversity view.'' First, Lehto, Juujärvi, Kooistra, and Pulkkinen (2003) found that Miyake's three-factor model provided the best fit of data from children aged 8 to 13. Second, Huizinga, Dolan, and Van der Molen (2006) employed CFA in a more developmental fashion by comparing the models of 7-, 11-, 15-, and 21-year-olds. They found partial support for the Miyake model as only the WM and shifting measures loaded onto latent variables, whereas the inhibition measures did not load onto a common latent variable (see also Miyake, 2009, for similar results with adults). Importantly, this model was consistent across the age groups, suggesting the stability of the EF construct across middle childhood, adolescence, and early adulthood. Together, these studies provide considerable evidence that Miyake's integrative model of interrelated, yet dissociable, EF components may be a suitable theoretical framework from which to examine EF development. However, these studies also suggest that the degree of unity and independence of the three components may change developmentally. This more complete picture of development and more nuanced version of the Miyake et al. model would be missed by focusing on only adults or a narrow age range on children, during the preschool years.

When a sample does include older school-age children and adolescents, methodological challenges can arise. First, to avoid ceiling effects researchers often use complex EF tasks that likely tap into multiple EFs, a problem of task impurity (Miyake et al., 2000). For complex tasks like the Wisconsin Card Sorting Test (WCST), the Tower of London (TOL), or Tower of Hanoi (TOH), task completion likely requires the coordination of multiple processes (e.g., Asato, Sweeney, & Luna, 2006; Huizinga et al., 2006; Miyake et al., 2000). However, for simplicity, researchers often classify tasks by a single cognitive construct. For example, the WCST and its child-appropriate version (DCCS) have been described as inhibition tasks by some and shifting tasks by others (Garon et al., 2008); the TOL and TOH have been described as either inhibition, WM, or planning tasks in various publications (e.g., Berg & Byrd, 2002; Huizinga et al., 2006; Welsh, Satterlee-Cartmell, & Stine, 1999).

Second, and very much relatedly, the tasks used across an age range often are not uniform. Tasks too difficult for the younger participants sometimes are only administered to the older ones, which makes comparisons across age groups difficult (e.g., Klenberg, Korkman, & Lahti-Nuuttila, 2001). Or very different tasks are used to assess a particular dimension for preschoolers and older children (e.g., Hughes, 1998; Welsh et al., 1991).

Keeping these issues in mind, we utilize Miyake's ''unity and diversity'' theoretical framework to focus on the ''foundational'' EFs—inhibition, information updating, and monitoring (WM), and shifting (Hughes, 1998; Huizinga et al., 2006; Lehto et al., 2003; Miyake et al., 2000)—in part because several frequently used cognitive tasks ostensibly tap into each dimension (Miyake et al., 2000). In using this framework to address EF development from early childhood through adolescence, we keep in mind current developmental theories of EF. In general, most of these theories depict EF development as involving an increasing ability to resolve conflict. They differ in whether this conflict is between rules that eventually become hierarchically organized (Zelazo et al., 2003), latent and active representations (e.g., habits vs. attention and WM; Munakata, 2001), or the current representation versus prepotent mental sets or behaviors (Diamond, 2006). Most also emphasize the role of changes in underlying neural networks. In particular, Posner and Rothbart (2007; see also Garon et al., 2008) propose that the development of the anterior attention system plays the major role in the resolution of conflict by regulating other brain networks. Posited developmental change is both qualitative (e.g., changing from simpler to more complex rule systems; Zelazo et al., 2003) and quantitative (e.g., strengthening active representations so that they override latent representations; Munakata, 2001).

We focus on those studies that most clearly address developmental issues—those that examine both preschoolers and school-age children, or school-age children and adolescents; address the order of acquisition of different aspects of EFs; or examine possible developmental processes. This approach permits us to detect developmental trajectories, sequences, and processes. Given a recent extensive review (Garon et al., 2008) of the large literature on the preschool age, only representative studies of this age are included.

Converging evidence and multiple levels of analysis are provided by studies using neuroscience techniques (e.g., functional magnetic resonance imaging [fMRI], event-related potential [ERP]) that assess the neural response underlying EF. It has been known for years that patients with PFC damage can have EF deficits yet normal IQ (e.g., Stuss & Benson, 1984). More recent thinking about this is that the PFC coordinates posterior cortical and subcortical brain activity via excitatory and inhibitory pathways (Casey, Amso, & Davidson, 2006; Shimamura, 2000). Moreover, PFC activity holds relevant information in WM (e.g., ''the cracker is hidden under the left cup'') and prevents distracting information from entering WM (Goldman-Rakic, 1987; Olson & Luciana, 2008; Shimamura, 2000).

We also know from structural imaging studies (e.g., using MRI) that PFC development, like brain development more generally, consists of both progressive (e.g., myelination, neuron proliferation, synaptogenesis) and regressive (e.g., cell death, synaptic pruning; Casey et al., 2006; O'Hare & Sowell, 2008) changes. The PFC matures later in adolescence as evidenced by further loss of gray matter (Gogtay et al., 2004; O'Hare & Sowell, 2008), unlike many other brain regions that mature earlier (e.g., regions involved in attention, motor and sensory processing, and speech and language development). During this time, progressive and regressive changes (largely myelination and synaptic pruning, respectively) occur concomitantly and are driven in part by the child's experiences—the result being ''efficient networks of neuronal connections'' (O'Hare & Sowell, 2008, p. 24).

Developmental neuroscience studies can enrich our understanding of EF development by determining how the neural correlates of behavior change over time. Changes in neural correlates (i.e., the neural response underlying task execution), in turn, can be interpreted in light of the known structural development of the brain and of the PFC in particular. Alternatively, changes in brain structure can be correlated with changes in task performance to determine the relevance of structural changes to EF maturation. In either case, we must remember that both progressive and regressive structural changes may influence how the neural response changes over time.

Foundational Executive Functions

Inhibition

Inhibition is considered foundational for EF (e.g., Miyake et al., 2000); however, most inhibition tasks are not pure measures of inhibition (Simpson & Riggs, 2005) nor do they tap into a single inhibitory process (Nigg, 2000). Garon et al. (2008) distinguished simple from complex response inhibition tasks based on whether WM also is needed. Simple response inhibition requires a minimal amount of WM, making it one of the purest forms of inhibition (Cragg & Nation, 2008). It shows its rudiments during infancy (see Garon et al., 2008), as when a child can delay eating a treat. Complex response inhibition also requires substantial WM by requiring that an arbitrary rule be held in mind or by requiring that the child inhibit one response (prepotent or not) and produce an alternative response. The day-night task assesses complex response inhibition by requiring the child to inhibit a prepotent verbal response (i.e., saying ''day'' upon viewing a picture of a sun) and activate an alternative verbal response (i.e., saying ''night'' upon viewing a sun; Gerstadt, Hong, & Diamond, 1994). Similarly, Carlson and Moses (2001), using factor analysis, distinguished delay tasks, which require withholding a propotent response, from conflict tasks, which require the child to make a response that conflicts with a prepotent response. Thus, the day-night task and Luria's hand game are considered conflict tasks (as well as complex response inhibition tasks) because they require the child to respond in a way conflicting with the natural response (i.e., associating a picture of the sun with night time and making a fist when shown fingers, respectively). Finally, Nigg (2000) distinguished several forms of inhibition that cover cognitive, behavioral, and emotional regulation.

Age differences. Garon et al. (2008) described rapid improvements in early childhood on a variety of complex response inhibition tasks (i.e., conflict tasks), such as the day-night task and Luria's hand game (see also Carlson & Moses, 2001; Hughes, 1998; Lehto & Uusitalo, 2006; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Despite their apparent similarities, different conflict tasks show different ages of mastery, perhaps indicating different cognitive demands. For Luria's hand game, which requires children to make a fist when shown a finger and vice versa, the most improvement typically occurs between ages 3 and 4 (Hughes, 1998); however, for the day-night task, 3- and 4-year-olds find it equally difficult (Carlson, 2005) and improvements may continue into middle childhood (Gerstadt et al., 1994). Furthermore, preschool children perform better on Luria's tapping task than the day-night task (Diamond & Taylor, 1996). Like Luria's hand game, the tapping task requires the inhibition and activation of hand motor responses, whereas the day-night task requires the inhibition and activation of verbal responses. In addition to different response modalities, Diamond and Taylor (1996) argue that the two tasks differ in the degree of response prepotency: There is a stronger tendency to say ''day'' when shown a sun than to mimic the motor movement of another person. However, evidence of a mirror neuron system that facilitates the imitation of hand gestures (e.g., Iacoboni & Dapretto, 2006) calls into question whether inhibiting the mimicking of hand movements is necessarily easier. In any case, preschoolers have some ability, although still immature and sensitive to task demands, to override a naturally prepotent response in favor of an alternative.

The Dimensional Change Card Sort (DCCS) is another complex response inhibition task used frequently with preschool children. Rather than requiring the inhibition of a naturally prepotent response, the DCCS creates a prepotent response during the preswitch phase that must later be inhibited. The child is shown a deck of cards that vary on two dimensions—shape (e.g., rabbit vs. battleship) and color (e.g., red vs. blue). During the preswitch phase, the child must sort the cards according to one dimension (e.g., color; ''If it's red, it goes here; if it's blue, it goes here''). In the postswitch phase, the child is asked to sort the cards by the other dimension (e.g., shape). Similar to other conflict tasks, reductions in perseveration occur from ages 3 through 4 (Carlson, 2005; Zelazo et al., 2003).

Manipulations of the standard DCCS provide insight into the developmental sequence of inhibition (e.g., Zelazo, 2006; Zelazo et al., 2003). In a No Conflict DCCS, the inhibitory demands are minimized, but the WM demands are maintained by presenting four nonoverlapping rules by which to sort the cards (e.g., sort by two colors at preswitch and by two shapes at postswitch). Here, 3- and 4-year-olds perform equally well, suggesting that the WM demands by themselves are not the cause of difficulty for the younger children in the standard DCCS (Zelazo et al., 2003). In an Advanced DCCS, a third sorting dimension is added; if there is a star on the card, the child should sort by color, but if there is not a star, the child should sort by shape. Five- and six-year-olds find this task difficult, showing a < 50% chance of passing (Carlson, 2005). Thus, as predicted by Zelazo's cognitive complexity and control (CCC) theory (Zelazo et al., 2003), the complexity of a task, defined in terms of the hierarchical structure of the child's rule system, is critical to task performance. The complexity of the child's rule system increases as the child integrates and embeds seemingly incompatible rules based on color and shape: ''If sorting by color, the blue one goes here; if sorting by shape, the rabbit goes here.'' An inability to integrate the rule systems causes perseverative errors; that is, the child will continue to sort the cards based on the initial dimension—color or shape. The Advanced DCCS requires the integration of another rule based on the presence of a star, whereas the No Conflict version does not require the integration of any rules—supporting the notion that the integration of rules is critical for inhibition development.

Findings of further improvement in inhibition after age 5 are mixed. In a rare study examining a wide age range, Klenberg et al. (2001) found improvement from ages 3 to 6 on the Statue task (maintaining a body position while the experimenter attempts to distract the child) and the Knock and Tap game (e.g., tap when the experimenter knocks and vice versa), but no further significant improvement through age 12. However, these tasks may have been too easy for the older children—perhaps because of the low prepotency of the response to be inhibited. A similar problem of finding appropriate tasks when using a wide age range is that the cognitive battery used in this study, the NEPSY, contains subtests not suitable for the youngest children. Thus, the 3- and 4-year-olds only completed 3 of the possible 12 subtests, making comparisons across the entire age range difficult. Still, in support of the conclusion that inhibition stabilizes by the early school years, Lehto et al. (2003) found no significant changes in inhibition between ages 8 and 13 on the TOL and Matching Familiar Figures tasks (although the inhibitory aspects of these tasks are not clearly identified, and see Steinberg et al., 2008, for continued improvements in an impulsivity measure of TOL performance through adolescence and early adulthood).

Other studies find further development after age 8. Interestingly, many of these studies have utilized computerized tasks such as the Go-No-Go task or the continuous performance task (CPT), both of which require a response to certain stimuli and inhibition of response to other stimuli (Brocki & Bohlin, 2004; Casey et al., 1997; Cragg & Nation, 2008; Johnstone et al., 2007; Jonkman, 2006; Jonkman, Lansbergen, & Stauder, 2003; Lamm, Zelazo, & Lewis, 2006). In the Go-No-Go computer task, the child must respond (by pressing a designated keyboard button) only to "go" stimuli (e.g., all letters except *X*) and inhibit response to the "no-go" stimulus (e.g., the letter *X*). The CPT adds a cue prior to go and no-go stimuli. For example, the child may be asked to respond to the letter *X*, but only when preceded by the letter *A*. Thus, responding to a no-go stimulus is considered a failure of inhibition ("commission errors"). Unlike conflict tasks, these tasks do not require the execution of an alternative response.

In one of these studies (Brocki & Bohlin, 2004), significant improvements occurred from ages 7 to 9 to 11 on behavioral measures that loaded onto a factor that the authors labeled a "disinhibition" factor (CPT disinhibition, CPT impulsivity, CPT inattentive impulsivity, and Go-No-Go commissions). Likewise, both Jonkman et al. (2003) and Casey et al. (1997) found significant decreases in commission errors on these tasks between age 9 and young adulthood (see also Jonkman, 2006; Klimkeit, Mattingley, Sheppard, Farrow, & Bradshaw, 2004). Cragg and Nation (2008) used a modified Go-No-Go task in which children had to depress a home key prior to pressing a target key. This modification allowed for partial commission errors (depressing of home key but not pressing target key in response to the no-go stimulus) in addition to traditional commission errors. In comparing 5- to 7-year-olds with 9- to 11-year-olds, only the partial commission measure was sensitive to developmental change, revealing that older children were able to inhibit a motor plan (releasing the home key and pressing the target key) at an earlier stage of execution than younger children. Johnstone et al. (2007) also found that the traditional no-go commission error was insensitive to change from ages 7 to 12. In contrast, they did find that performance on a Stop-Signal task, during which the child must inhibit a currently activated response, was sensitive to change across this age span. Thus, these two studies suggest that the stage of execution is a factor in inhibition difficulty: Terminating an already executed response appears to be more difficult than inhibiting a response that has yet to be executed or is in an earlier stage of execution.

In a computerized anti-saccade task (Fischer, Biscaldi, & Gezeck, 1997; Munoz, Broughton, Goldring, & Armstrong, 1998), children fixate on a central target. The target is turned off and a peripheral cue is turned on. Children are told not to look at the target but instead to look to the opposite side. Thus, any initial glance toward the cue is an inhibitory failure. Dramatic improvement in both reaction time and accuracy during the grade school years is followed by slower improvement during early adolescence. Similarly, Williams, Ponesse, Schachar, Logan, and Tannock (1999) found improvement up through age 12 on a "stop-signal reaction time" task, involving inhibition of a response (key press) to stimuli when a tone sounds. Moreover, Huizinga et al. (2006) found continued improvement in both reaction time and accuracy measures on the Stop-Signal task and Eriksen Flankers task until age 15 and on a Stroop-like task (inhibiting saying a color word in order to state its conflicting font color) until age 21. Finally, adults, more than adolescents, appeared aware of making an inhibition error as they momentarily slowed their response for the next trial in order to prevent further error (Hogan, Vargha-Khadem, Kirkham, & Baldeweg, 2005), which suggests the contributions of metacognitive development even after adolescence.

The fact that several studies have found improved performance beyond early childhood on ostensibly simple response inhibition tasks (e.g., anti-saccade task)—or at least on tasks (e.g., Go-No-Go task) simpler than conflict tasks—challenges the notion that performance on these tasks matures early on. For example, the computerized Go-No-Go task should be easier than Luria's hand game because it requires only response inhibition of a prepotent response rather than response inhibition followed by the execution of an alternative response. One explanation for this involves how task performance is measured. Computerized tasks contain multiple trials and measure reaction time very precisely (to the millisecond), thus increasing sensitivity to subtle changes. Alternatively, since many of these studies with older children have utilized computer-based tasks, perhaps some of the age-related improvements are related to computer-specific abilities (e.g., more efficient use of keyboard or mouse).

In summary, the first leap in attaining inhibition appears in the preschool years. By age 4, children show signs of successful performance on both simple (i.e., pure response inhibition) and complex inhibition tasks (i.e., response inhibition plus alternative response). By the same age children can operate on a bidimensional card by one dimension and then inhibit that to use the second dimension (i.e., successfully perform DCCS). Inhibition continues to improve, particularly from age 5 to 8 (Romine & Reynolds, 2005) and particularly for tasks that combine inhibition and WM (Carlson, 2005; Gerstadt et al., 1994), but also at later ages, especially on computerized tasks. Unlike the early improvements, these are unlikely to be fundamental changes in cognition (e.g., like preschoolers' acquisition of the rule-formation ability needed to perform the DCCS). Instead, refinements seem to involve quantitative improvements in accuracy, perhaps due to an increasing efficiency to override prepotent responses. Thus, it may be that inhibition tasks have varying sensitivities, with some being sensitive to the conceptual gains in early childhood and others being sensitive to the refinements in strength of the relevant cognitive skills or the generality of application in later childhood and even adolescence. These sensitivities seem to be determined by a number of factors including how performance is measured, the response modality, the strength of the response bias, the stage of response execution, and the degree of simultaneous WM demands imposed by the task.

Evidence from neuroscience. One method to examine the neural response underlying response inhibition is to measure the brain's electrical activity via electroencephalogram (EEG). Measuring during infancy, early childhood, and middle childhood, one longitudinal EEG study (Bell, Wolfe, & Adkins, 2007) reported qualitative change in brain activity underlying complex response inhibition—measured by WM and inhibitory control tasks. At 8 months of age, correct performance on the A-not-B task was associated with increased global cortical activity, whereas at age 4½, day-night performance was associated only with increased medial-frontal activity. By age 8, this activity became even more focused in the right frontal scalp regions during completion of the WCST. This shift from global to localized activity during task completion may signal the growing efficiency of the brain and the growing functionality of the PFC for complex response inhibition.

EEG studies with older children indicate continued localization of brain activity from middle childhood to adulthood. Looking closely at frontal activity, Jonkman (2006) found a linear increase in no-go P3 amplitude (a positive wave 300–500 ms after stimulus presentation) across development: Children aged 6–7 showed no P3 activity, children aged 9–10 showed limited P3 activity in frontal-central electrodes, and young adults showed broad P3 activity across frontal and frontal-central electrodes (see also Jonkman et al., 2003). P3 activity has been interpreted to index the allocation of attentional resources during stimulus engagement, with greater activity representing greater resource allocation (Polich, 1987). In contrast, the no-go N2 response (a negative wave 150–400 ms after stimulus presentation) decreased across the same age span, with the largest decrease in amplitude occurring between ages 6 and 10. The N2 response is thought to indicate conflict monitoring, needed when the prepotent response conflicts with the task-required response. The decrease in the N2 response may indicate that compensatory mechanisms are present in younger children and that the task induces less conflict with increasing age (Jonkman, 2006).

Similar findings come from Lamm et al. (2006) who reported decreases in frontal N2 amplitudes from ages 7 to 17 on no-go trials, which they attributed to increasing neural efficiency that may result from the regressive neural changes described above (e.g., synaptic pruning). As evidence that frontal N2 activity reflects conflict monitoring and inhibition, performance on tasks requiring conflict monitoring and inhibition (the Stroop task and Iowa Gambling Task) predicted decreases in frontal N2 amplitudes beyond that predicted by age. As further support, Johnstone et al. (2007) observed stronger frontal N2 activity in response to no-go stimuli than go stimuli. However, they found no change in the ability to inhibit no-go responses across the ages of 7–12. Moreover, what age-related changes in brain activity the authors did observe were found in nonfrontal brain regions—including decreases in central and parietal N2 amplitude—suggestive more of refinements in stimulus processing than in the inhibition response per se. All together, these EEG studies provide evidence that inhibition development is paralleled both by increases and decreases in neural activity, perhaps indicative of progressive and regressive neural change.

Researchers also have used neuroimaging assessments, such as fMRI, to document the increasing efficiency of the neural response underlying response inhibition. On the Go-No-Go task, Casey et al. (1997) observed no difference in the *location* of activation in the PFC between children ($M_{\text{age}} = 9.92$) and young adults but did measure *greater volume* of activation in children during the no-go condition, particularly within the dorsal and lateral PFC. Moreover, a longitudinal imaging study (Durston et al., 2006) reported that activation within the PFC only increased from ages 9 to 11 for specific prefrontal regions (ventral PFC) correlated with task performance. In contrast, activity within prefrontal regions uncorrelated with task performance (dorsolateral PFC) decreased with age. Similar to the explanation offered by Lamm et al. (2006), Casey et al. and Durston et al. (2006) suggested that the greater activation in children may correspond to inefficiency in the inhibition mechanism and that neural development is characterized by increased localization of activity to brain regions directly linked to the behavioral response and decreases in activity in supplementary brain regions.

One structural imaging technique, diffusion tensor imaging (DTI), measures the myelination of axons. Myelinated axons transmit signals more efficiently and rapidly than unmyelinated axons. Liston et al. (2006) measured myelination in participants aged 7–31, who also performed a Go-No-Go task and found that myelination of projections from the PFC to the striatum increased with age and correlated with no-go task performance. Importantly, myelination of corticospinal axons also increased with age but did not correlate with task performance. Thus, the increasing myelination of frontostriatal connections, in particular, across childhood and adolescence seems to underlie the maturation of response inhibition.

A synthesis of these neuroscience studies does not suggest a one-to-one correspondence between changes in brain activity and changes in task performance. Instead, task performance often changes subtly, if at all (e.g., Johnstone et al., 2007), whereas the pattern of neural activity may change dramatically. It seems that school-age children can successfully complete response inhibition tasks (with concurrent WM requirements or not) but, in doing so, enlist a more global pattern of activation than they will later on. With development comes localized and efficient activation in specific PFC regions (e.g., ventral PFC) pertinent to task completion. These dramatic changes in neural activity may translate into only subtle improvements in response inhibition, such as greater efficiency and less effort.

Developmental issues. Several developmental issues about inhibition need to be examined further. First, by what process does a child move from one level (e.g., simple inhibition of a response) to another level (e.g., the development of a more complex rule that controls two responses)? Does the former contribute to the development of the latter? Would examining the transition period yield clues to the qualitative and quantitative aspects of this change? Second, what is the form of the developmental trajectory from age 3 to adolescence? Is development linear? This question is difficult to answer because of the difficulty of finding appropriate tasks for a large age span. The dramatic improvement from ages 3 to 5 appears to be followed by less dramatic change from 5 to 8 and even less change after age 8 (although brain maturation continues). However, the tasks used with younger children likely have different inhibitory requirements than those used with older children. In a meta-analysis of EF studies from age 5 to adulthood, Romine and Reynolds (2005) found the greatest advancements in inhibition of prepotent responses from ages 5 to 8. A meta-analysis that includes younger children is greatly needed to clarify the rate of change from the preschool to grade school years.

Relevant to this apparent developmental trajectory is the problem of task impurity. Inhibition appears not to be a uniform construct. For example, interference control, cognitive inhibition, and motor inhibition may be distinct processes tapped by different tasks and develop at different rates across childhood (Nigg, 2000). Furthermore, classic tasks of inhibition utilize cognitive components in addition to inhibition. As already mentioned, many tasks also have significant WM requirements. For another example, performance on the Stroop task may depend not only on the ability to inhibit reading a color word in order to read its actual color but also on the child's level of reading automaticity (Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004). As a result, performance, measured in terms of reading errors, does not improve in a linear fashion, but rather forms a quadratic relation: There is an initial increase in reading errors from ages 6 to 10, followed by a dramatic decrease in errors through age 17. This suggests that as word reading becomes more and more automatic from ages 6 to 10, inhibition of that process to say the color becomes more difficult, which negatively affects reading accuracy. Afterward, the inhibition mechanism needed may be mature enough to compensate for this reading automaticity. Thus, this developmental trajectory may have as much to do with developing reading automaticity as with developing inhibition. A similar careful analysis of other inhibition tasks might reveal similar complexities regarding the question: ''What develops?''

Adding further complexity, social factors likely influence inhibition performance. For example, there is evidence for a developmental regression in self-regulation in early adolescence (Anderson, 2002), particularly as evidenced by increased risky behavior during this time (Steinberg, 2007). One perspective suggests that the interaction of a mature socioemotional network (involving mainly limbic structures) and an immature EF-related control network within a context of increased peer pressure leads to adolescents' increased risk taking (Steinberg, 2007). This perspective suggests that outward regressions may occur not so much because there are actual regressions in inhibition per se but because exposure to emotionally laden and risky situations increases at this point in development when inhibition is still immature.

Summary. Regarding the first component in the Miyake et al. (2000) model, cognitive, behavioral, and brain assessments generally show rapid early improvements in inhibition followed by slower improvements through adolescence, along with greater brain localization throughout childhood and adolescence. Mechanisms of development may include brain maturation, increased ability to handle task complexity, increased ability to use rules, and emerging metacognition.

Working Memory

Like inhibition, WM research has been complicated by various definitions of the construct, as well as (and perhaps resulting in) the use of differing assessment tasks. In general, WM involves the ability to maintain and manipulate information over brief periods of time without reliance on external aids or cues (Alloway, Gathercole, & Pickering, 2006; Goldman-Rakic, 1987; Huizinga et al., 2006). Neuroimaging research suggests that WM tasks vary by how much the task elicits PFC activity; consequently, researchers suggest that WM tasks vary in the degree to which they require ''executive control'' (Luciana et al., 2005). That is, more complicated WM tasks that require the maintenance and manipulation of information in order to direct behavior toward future goals (e.g., backward digit span, delayed-response tasks, self-ordered searches) ostensibly rely on more executive involvement (and consequently more PFC activity; D'Esposito & Postle, 1999) than simpler WM tasks that require only the maintenance of information (e.g., forward digit span). Accordingly, the rate and form of WM development will hinge upon the degree to which a task requires executive processes.

Gathercole, Pickering, Ambridge, and Wearing (2004) invoked the classic WM model of Baddeley and Hitch (1974) to distinguish between executive WM tasks and online storage tasks. According to Gathercole et al., executive WM requires that either the verbal storage system (i.e., the phonological loop) or the visuospatial storage system (i.e., visuospatial sketchpad) work in concert with a coordinating central executive. Simple WM tasks require little input from the central executive but rely solely on either the phonological loop (e.g., forward digit recall) or visuospatial sketchpad (e.g., visual pattern recall). Complex tasks (e.g., backward digit recall), on the other hand, may require that multiple WM tasks be performed concurrently, and therefore the central executive must coordinate those processes. This tripartite model was supported by a CFA of children aged 6–15 on a battery of verbal WM, visuospatial WM, and executive WM tasks, which suggests the differentiation of the WM subsystems by early grade school.

Age differences. A variety of tasks document improvement in WM during the preschool years (see Garon et al., 2008). Gathercole et al. (2004)

reported that by age 6, the executive component of WM is sufficiently developed to be used during complex tasks that require coordination of WM subcomponents. In addition, the same researchers found that simple and complex WM tasks had similar developmental trajectories—a linear increase from ages 4 to 14 and a leveling off between ages 14 and 15 across nearly all tasks examined. Luciana et al. (2005), drawing on a battery of nonverbal tasks, found that the developmental course of WM depends on the complexity and, thus, the executive demands, of the task, with less demanding tasks being mastered earlier in development. Their battery of nonverbal tasks ranged from a nonverbal face recognition task (lowest executive demand) to a spatial self-ordered search (highest executive demand). The former simply required the child to maintain a facial representation over a delay in order to discriminate a previously viewed face from a novel one, and performance was unchanged between ages 9 and 20. The latter task required the child to search varying locations on a computer screen for hidden tokens, to remember locations where a token was found, and to strategically explore other locations, and performance continued to improve until age 16.

Extending the work of Luciana et al. (2005), Conklin et al. (2007) used a battery of verbal and spatial WM tasks across the ages of 9–17. Across age, children tended to perform better on the spatial versus verbal WM tasks, even though the tasks were thought to tap similar cognitive processes (e.g., strategic self-organization). Still, the developmental trajectories were similar for verbal and visuo-spatial WM tasks, and instead, differed based on task complexity. In agreement with the conclusions of Luciana et al., Conklin et al. suggest that the age of mastery depends more on the degree of *processing* (with more complex tasks requiring a greater degree of processing) rather than the *content* to be processed (e.g., verbal vs. visuospatial material).

One problem, although, with this use of different tasks to manipulate the executive demands on WM is that it is difficult to ensure that non-EF processes are equivalent across tasks and that they do not influence age differences in performance. Luciana and Nelson (1998) addressed this important issue by employing only a self-ordered search task, which varied over trials in the executive demands, based on the number of locations (2–8) the child may search for tokens. By increasing the number of search locations, greater demands are placed on the executive components of WM to search strategically and to avoid previously searched locations by continually updating WM. Similar to the findings of Luciana et al. (2005), for the least demanding condition, performance was equivalent among children aged 4–8, adolescents, and young adults. However, as the number of search locations increased, age differences emerged. For three locations, performance maturity was reached at age 6, for four locations, maturity was not reached until adolescence, and for six and eight locations, improvements continued until adulthood. Thus, the development of executive WM occurs gradually with continued refinement through adolescence, especially for tasks that require the maintenance and manipulation of multiple items.

Evidence from neuroscience. In accord with the behavioral results, fMRI evidence points to a protracted developmental course leading to localized activity within the PFC during WM functioning. Kwon, Reiss, and Menon (2002) reported a quantitative linear increase in activity within a frontoparietal network, including ventral and dorsal regions of the PFC, from ages 7 to 22 while performing a visuospatial WM task (*n*-back task). The authors note that the increased activity within right-lateralized dorsal PFC likely subserves the maturation of visuospatial attention and executive processes, whereas the increased activity within a left-lateralized frontoparietal network subserves the maturation of a phonological rehearsal system. Increases in this neural activity were related more to age than to task performance (accuracy and RT). Thus, with a dramatic and prolonged increase in specialization of the WM neural circuitry through childhood and adolescence comes only limited overt task improvement, especially during adolescence.

Scherf, Sweeney, and Luna (2006) reported both qualitative changes (location of activation) and quantitative changes (amount of activation) in the neural response underlying visuospatial WM from childhood ($M_{\text{age}} = 11.2$) through adulthood ($M_{\text{age}} = 29.5$). During childhood, activation occurred in qualitatively different premotor regions and also in the lateral cerebellum, which was absent in later development. Also during childhood, there was quantitatively greater activity in ventromedial regions, including the thalamus and basal ganglia. Adolescence ($M_{\text{age}} = 15.7$) brought a shift in activity to frontal regions, including the right dorsolateral PFC. Finally, from adolescence to adulthood, activity became more localized and lateralized as left dorsolateral PFC activation increased and right dorsolateral PFC activation decreased. Moreover, activity increased substantially in the anterior cingulate (described as qualitative change by the authors). The end result of these changes (quantitative and qualitative, progressive and regressive) from childhood to adulthood is a functionally specialized visuospatial WM neural network (see Klingberg, Forssberg, & Westerberg, 2002, for similar changes from ages 9 to 18 on a visuospatial WM task).

Increased WM also correlates with structural neural indices, such as the maturation of white matter (i.e., the myelination of neuronal axons), measured by DTI. Nagy, Westerberg, and Klingberg (2004) reported that the development of visuospatial WM from ages 8 to 18 correlated with myelination in regions primarily of the frontal lobe close to the parietal lobe (superior and inferior left frontal lobe), whereas the development of reading ability correlated with myelination in the left temporal lobe. Thus, during late childhood and adolescence, the maturation of specific cognitive functions is linked to the maturation of specific neural circuits, rather than to global brain maturation.

Developmental issues. One developmental issue concerns the developmental relation between WM and inhibition. It seems that many inhibitory tasks, particularly complex ones, also place demands on WM (Bell et al., 2007; Garon et al., 2008; Simpson & Riggs, 2005), and the combination of the two within a single task poses significant difficulty for young children (e.g., Carlson, 2005). For example, in the day-night task the child must maintain the rule in WM. However, research also suggests their independence. If the inhibitory component is eliminated on this task, the difference between 3.5- and 5-year-olds nearly disappears (Simpson & Riggs, 2005), which suggests that it is inhibition, not WM, that causes the age difference. (It is also possible that eliminating the WM component also would reduce the age differences, which would suggest that the interaction of the two components is important.) Similar outcomes were found with a Stroop-like task, a CPT task, and a start–stop task (Beveridge, Jarrold, & Pettit, 2002), as well as the DCCS (Zelazo et al., 2003). Thus, relevant to the Miyake et al. (2000) model, it appears that WM and inhibition are largely separate constructs (for a counterargument, see Bell et al., 2007; Davidson et al., 2006). That said, many tasks described as either WM or inhibition tasks likely place demands on both types of processes (e.g., the day-night task), and therefore, it is difficult to obtain a pure measure of one or the other.

Summary. Performance on complex WM tasks (i.e., those tasks requiring a greater degree of processing such as the maintenance and manipulation of information) improves at least through adolescence. Like the development of the neural circuitry subserving response inhibition, the development of the WM circuitry involves progressive and regressive changes, resulting in a localized pattern of activity within a frontoparietal network, including the DL-PFC. Unlike the trajectory of inhibition development that shows large improvements during the preschool years followed by more modest, linear improvements through adolescence, most of the evidence suggests that the trajectory of WM development is linear from preschool through adolescence.

Shifting

The third core EF is the ability to shift between mental states, rule sets, or tasks (Miyake et al., 2000). There appears to be substantial need for inhibition and WM processes for shifting. Reminiscent of the cognitive processes associated with inhibition tasks (e.g., DCCS), Miyake et al. (2000) suggest that shifting may "involve the ability to perform a new operation in the face of proactive interference or negative priming" (p. 56). It would seem, then, that the ability to inhibit previously activated mental sets would be important for successful shifting and that perseverative errors (i.e., continued responding based on the previous mental set) would indicate shifting failures (Anderson, 2002). The typical distinction between tasks deemed "inhibition tasks" and those deemed "shifting tasks" is that the latter typically rely on switching between two or more mental sets—with each set possibly containing several task rules—rather than the inhibition of a single response (Crone, Somsen, Zanolie, & Van der Molen, 2006). Moreover, in tasks of inhibition, the rules are usually explicitly expressed rather than implied through either negative or positive feedback. Thus, the DCCS with its explicit indication of set change is often categorized as an inhibition task (but see Garon et al., 2008, who classify the DCCS as a shifting task). Shifting tasks also place demands on WM by requiring the maintenance and updating of that mental set based on feedback.

Age differences. The ability to shift improves with age (Anderson, 2002; Cepeda, Kramer, & Gonzales de Sather, 2001; Crone, 2007; Crone et al., 2006; Garon et al., 2008; Somsen, 2007). Preschoolers, aged 3–4, can successfully shift between two simple response sets in which the rules are placed in a story context (Hughes, 1998) or when demands on inhibition are reduced (Rennie, Bull, &

Diamond, 2004). For example, in a simplified version of the WCST (Hughes, 1998), preschoolers can determine what a teddy bear's favorite shape is, based on feedback, and then after a set shift they can decide what a second teddy bear's favorite color is, based on differing feedback.

As previously mentioned, Senn et al. (2004) reported that whereas inhibition and WM were interrelated and predicted complex task performance (i.e., TOH), shifting was unrelated to inhibition, WM, or TOH performance in preschoolers. Accordingly, the authors suggest that shifting may not be differentiated from WM and inhibition—and therefore, is less developed—at this age. This is sensible given that inhibition and WM processes seem to be prerequisite processes for successful shifting. As Garon et al. (2008) noted, before children can successfully shift between response sets, they must be able to maintain a response set in WM and then be able to inhibit the activation of a response set in order to activate an alternative one.

More complex tasks show further development in older children and adolescents. Luciana and Nelson (1998) utilized a set-shifting task that progressed through nine stages of increasing difficulty and complexity (the intradimensional–extradimensional set-shifting task from the CANTAB). This is a fruitful strategy as it shows exactly how complex a set children can handle at each age, thus permitting comparisons across a wide age range. This task required children to respond correctly, based on previous feedback, to either lines or shapes presented on a computer. Children had to attend to feedback, infer the correct rule at that moment, and respond accordingly. At set points, reinforcement switched such that the correct response (e.g., shapes) switched to the opposite of what was previously correct (e.g., lines). The main improvement occurred from ages 5 to 6, at Stage 7, in which the rule did not switch (i.e., between lines and shapes), but the examples of lines and shapes did change. For successful completion of this stage, children needed to utilize feedback from previous stages to shift their response to new examples of either lines or shapes. With increasing age up through young adulthood, there was a steady increase in the proportion of subjects who completed all nine stages of the task, indicating that shifting ability continues to improve over many years.

Huizinga et al. (2006) investigated set shifting on three computerized tasks in which a cue signaled to which dimension the child should respond. Sporadically, the cue would switch, indicating a set shift. ''Shift cost'' was the difference in response time between shift trials and nonshift trials. The shift cost for the 7- and 11-year-olds was significantly greater than for the 15-year-olds, who did not differ from the young adults ($M_{\text{age}} = 20.8$). Thus, shifting reached adult-like levels around age 15. Similarly, Davidson et al. (2006) found improvement from age 4 through adolescence. Interestingly, they found different developmental trajectories for the switch cost for accuracy versus reaction time. Whereas the shift cost to accuracy diminished through early adolescence, the switch cost to reaction time increased until adulthood. This speed–accuracy trade-off indicates that with increasing age, participants were more likely to slow down their responses on shift trials to ensure that they were responding accurately. Thus, improved metacognition—knowing that slowing helps performance and being able to detect when it is advantageous to do so—may be one mechanism of developing accurate set shifting. This design, examining developmental curves on two or more measures, is a model for how future research could tease apart aspects of an EF component and tell a more nuanced developmental story.

Evidence from neuroscience. In further support of the role of metacognition, especially monitoring and changing one's own performance, Crone et al. (2006) measured heart rate changes during a task shifting paradigm, in which the child ''opened'' doors on a computer screen in order to help a computerized donkey find its way home. Heart rate slowing following negative feedback (indicated by a negative sign after opening an incorrect door) would indicate a realization of an error and the evaluation of the current set rules. For ages 8–18, heart rate slowed to a similar degree following the unexpected feedback that occurs immediately after a task shift. The difference between the age groups occurred only in errors that continued after receiving feedback for a task shift. Following such an error, the 8- to 10-year-olds did not show heart rate slowing as much as the 12- to 14-year-olds or the 16- to 18-year-olds. Thus, although younger children could detect a task shift as well as older ones, they did not detect performance errors after the shift as well. Somsen (2007) reported that on a computerized WCST an increase in attention to feedback about errors predicted performance in adolescents, but not younger children, which again supports the role of metacognition, particularly being able to use feedback to change one's behavior.

fMRI analyses implicate neural activity within multiple regions of the PFC and elsewhere as shifting develops. Rubia et al. (2006) reported increased activation in inferior frontal, parietal, and anterior cingulate regions, but decreased DL-PFC activation across adolescence during shifting. They proposed that the increased activity in the anterior cingulate cortex (ACC) reflects the maturation of conflict monitoring processes, whereas the increased DL-PFC activation in younger participants reflects compensatory neural activity—quite similar to the explanation offered by Casey et al. (1997) regarding Go-No-Go performance.

Also focusing on conflict monitoring, Crone (2007) suggested that other regions may underlie the growing ability to monitor and change one's performance. During development, adult levels of processing feedback about performance on a shift task are reached first for the medial PFC (important for violations of processing expectations), and then for the left dorsal PFC (important for hypothesis testing and seeing the need for adjustment of behavior). The first development occurs between ages 8 and 10 and adolescence and the second between adolescence and adulthood. Thus, because cognitive shifting requires the child to switch between multiple response sets based on feedback, neural networks involving the ACC and regions of the PFC that are responsible for monitoring and detecting conflict (e.g., performing a response and receiving negative feedback) seem to be critical to successful shifting.

Developmental issues. A useful approach for pinpointing exactly *what* changes with age on a shifting task is to break such a task down into component processes and examine neural correlates of age changes in these processes. A recent interesting study (Morton, Bosma, & Ansari, 2009) examined separately two processes in the DCCS task: (a) rule switching (e.g., switching from color- or shape-based sorting) and (b) detection and resolution of conflict when stimuli can legitimately be sorted in two ways (color or shape). Rule switching showed fMRI activity in the lateral PFC and posterior parietal cortex in both children and adults. Most interesting, however, was an interaction of rule switching and age in these regions. Rule switching modulated activity in the left posterior parietal cortex and right middle frontal gyrus in adults but not in children. Thus, the networks involved in switching presented a qualitative developmental change from childhood to adulthood. Although the age-related results were less clear for conflict processing, they showed that conflict processing and rule switching are separable processes that, when considered separately, can give a more fine-grained analysis of age-related neural and behavioral changes on the DCCS task.

Age-related improvements in shifting also occur through the development of processes other than shifting per se. For example, the ability to generalize a rule set to a novel set of stimuli facilitates shifting performance (Luciana & Nelson, 1998). The abilities to maintain the new rule set and to detect performance errors after a successful shift also are important (Crone et al., 2006). Finally, the development of metacognitive strategies, such as slowing down responses to preserve a high level of accuracy, enhances accurate shifting (Crone et al., 2006; Davidson et al., 2006).

Summary. The ability to successfully shift between task sets follows a protracted development through adolescence. It appears that preschool-aged children can handle shifts between simple task sets and later can handle unexpected shifts between increasingly complex task sets. Both behavioral and physiological measures indicate that during adolescence, monitoring of one's errors is evident, and by middle adolescence, task switching on these complex shift paradigms typically reaches adult-like levels. Because of greater need for multiple cognitive processes, mature shifting likely involves a network of activity in many PFC regions.

Conclusions About Development and Directions for Future Research

Developmental Trajectories

It is clear that previous reviews of EF in children, mainly focused on preschoolers, leave out much of the story of the development of EF and limit the search for sequences and mechanisms of development. There is substantial further development in all components after age 5, and even through adolescence (see Best et al., 2009). Including broader age ranges provides information about developmental trajectories. Results are inconsistent, but inhibition appears to show particularly striking improvement during the preschool years and less change later on. WM shows more gradual linear improvement throughout development, as does shifting. These different trajectories provide support for the Miyake et al. (2000) position that the three components are somewhat diverse. However, the differing trajectories extend the Miyake et al. model by suggesting that the degree of unity or diversity of EF varies from age to age.

The evidence suggests both quantitative and qualitative EF development. Much of the change appears to be quantitative and gradual, although the change may be more rapid in the early years. In many relevant brain regions, activity decreases with age, perhaps reflecting the growing efficiency of the neural response. Some change appears to be qualitative, suggesting changes in brain organization as the site of brain activity shifts during development (e.g., Scherf et al., 2006). In other regions, activity seems to increase—for example, in the ventral PFC during response inhibition and in the DL-PFC during WM. As Olson and Luciana (2008) note, the discrepant behavioral trajectories along with this segregation of activity tentatively suggest that different regions of the PFC support different EFs. Thus, regional differences in the course of neural development may be responsible for different developmental trajectories of response inhibition, WM, and shifting. The emergence of metacognition may also bring qualitative change when children learn to use feedback about errors to change their approach to the task.

It is important to note that despite evidence for the functional differentiation of the PFC, there also seems to be activation of common regions during the completion of distinct EF tasks. This may in part be due to EF task impurity—notably the fact that it is difficult to tease apart WM and inhibitory processes (Roberts & Pennington, 1996) and that shifting likely builds on WM and inhibitory processes (Garon et al., 2008). Additionally, it may indicate the common processes that underlie the various EF components, that is, the "unity" aspect of the "unity and diversity" of EF (Miyake et al., 2000). Exactly how EF is instantiated in the brain, including to what degree the PFC is functionally differentiated, continues to be a debated issue (Olson & Luciana, 2008).

Several task-related factors also influence the observed developmental trajectory of a latent EF component. One is the degree of task complexity, with better success expected on simpler tasks. Task complexity can be manipulated in several ways, but notably by increasing the degree of response prepotency in inhibition tasks (Diamond & Taylor, 1996) or by increasing the degree of "working" with information in WM tasks (Luciana et al., 2005). Thus, a simple response inhibition task should be easier than the Stroop task, and a WM task that requires online storage should be easier than one that requires information manipulation. A second is how performance is measured (e.g., partial vs. full commission error on the Go-No-Go task; Cragg & Nation, 2008). A third is the response modality (e.g., bodily vs. verbal response; Diamond & Taylor, 1996). Thus, researchers need to consider carefully how these task factors may shape the developmental trajectory of constructs ostensibly measured by such tasks. Moreover, the influence of these task factors suggests caution in drawing strong conclusions about the development of EF.

Mechanisms of Development

Research needs to move beyond a focus on description—the ages at which the EF components emerge, show rapid development, and reach maturity—and address mechanisms of development. How do children move from early to later levels of competence within an EF component, for example, inhibition? Does the early phase of development of one component facilitate the development of other components? The developmental relations between components of EF could be examined by research designs that include assessments of several components of EF together in the same study, at different ages. The comparison of trajectories of several EF components can suggest mechanisms of development. For example, the early rapid improvement in inhibition may contribute to the later developments of shifting and planning.

Promising designs. Several designs used by studies reviewed here seem particularly promising for examining mechanisms of development within or between components: (a) Use meta-analyses (Romine & Reynolds, 2005) to examine the effects of moderating variables at different ages by including a larger age range. (b) Compare the developmental trajectories for two or more aspects of performance (e.g., speed and accuracy, Davidson et al., 2006; Baker, Segalowitz, & Ferlisi, 2001; number of moves, amount of planning time prior to making the first move, and the proportion of perfect solutions on the TOL task, Huizinga et al., 2006) for clues as to whether one aspect influences another. (c) Look for correlations between a measure of neural activity and EF performance (e.g., Morton et al., 2009). Brain maturation could permit, and thus serve as a developmental mechanism for, more advanced EF behavior, or the latter behaviors could cause changes in neural networks.

Another promising, yet rarely used, approach is to focus on the transition phase from one developmental level to the next. An example is the change from the simple rule necessary for inhibition on simple conflict tasks to the more complex rules necessary for inhibition on the DCCS task. The microgenetic method is particularly useful in this regard. In this method, children have multiple trials on the same task or similar tasks, typically in several sessions over several weeks, and trial-to-trial changes are examined. Using this method, McNamara, DeLucca, and Berg (2007) tracked detailed change in the type of strategy used with increased experience on the TOL task. Potentially a microgenetic design could show that rapid change in one component is subsequently followed by rapid change in another component, suggesting a possible causal relation.

Microgenetic studies could be particularly useful for examining what happens at the point of shifting on shift tasks, given that several studies have observed changes that suggest metacognitive processes at work on the trials after the shift (e.g., Crone, 2007; Crone et al., 2006). In a task similar to shift tasks (DeMarie-Dreblow & Miller, 1988) when young children have to shift from one category (e.g., animals) to another category (household items) on a selective memory task, their previously effective strategy of attending to only relevant items transferred successfully to the new category but temporarily became ineffective at facilitating recall of the relevant items in this newly relevant category. Such fine-grained assessments of microgenetic changes in behavior would be particularly powerful if combined with neuroimaging assessments that track changes in brain activity at the point of behavior change.

Training studies also are useful for examining possible mechanisms of development. These can look at short-term change by (a) training one EF component and then assessing any immediate changes in other components, or (b) providing particular experiences, such as metacognitive instruction, and observing any facilitation of EF.

Training studies also can examine mechanisms over longer periods of time. The most powerful assessments of developmental mechanisms would track changes in both cognitive performance and brain organization, thus providing a multilevel assessment of influences. Such studies are rare. One such line of investigation showed both improved EFs and change in fMRI patterns of brain activity associated with EF after a 3-month after-school high-intensity exercise program for ages 7–11 (Davis et al., 2007, in press). In comparison to a control group, the exercise group had increased bilateral PFC activity and reduced bilateral posterior parietal cortex activity during an inhibition task. Another study (Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005) reported evidence that 1 week of training on a computerized program centered on executive attention led to a more mature EEG response to an inhibition task (modified flanker task) and to small, although often insignificant, behavioral improvements on the inhibition task and on a generalized intelligence test.

Finally, a powerful approach to identifying developmental mechanisms underlying changes in EF would be to study whether the cognitive, biological, and social correlates of EF change from one age to another (see Best et al., 2009, for a discussion of the social correlates of EF). This is particularly important because different processes may contribute to the development of EF at different ages.

Developmental sequences. Another good starting point to search for possible mechanisms of development is to examine developmental sequences underlying the emergence of each EF component (for a description of types of developmental sequences, see Flavell, 1972). Observed developmental sequences can suggest how early cognitive skills might be related to later ones. For example, does the development of a more advanced inhibitory ability *supplement* earlier inhibitory ability (i.e., an "addition" sequence), for example, by strengthening it, adding the ability to select an alternative response to the ability to inhibit the prepotent response, or adding metacognitive skills? Or does a more advanced, integrated inhibition skill *replace* earlier forms of inhibition, as suggested by the evidence of brain reorganization (e.g., the loss of compensatory activation with development) correlated with more advanced inhibition? In this qualitative change, the old brain organization associated with simple inhibition may no longer exist. The microgenetic design mentioned above could address supplementing versus replacing by showing whether Skill A continues to be used even after Skill B emerges.

A similar analysis of sequences and possible mechanisms could also address how the various EF components are related developmentally. That is, some EF components may facilitate the development of other EF components. For example, as WM and inhibition seem to develop ahead of shifting (Davidson et al., 2006), perhaps a certain level of WM and inhibition has to be developed before children can use them toward the development of shifting behaviors (Garon et al., 2008). Two EF components may even be mutually facilitative, as each bootstraps the development of the other in a back-and-forth fashion. Garon et al. (2008) suggested that the emergence of the three components in the first 3 years of life may be followed by an integrative period in which they become coordinated. In short, a greater emphasis on detecting possible developmental sequences will allow for a clearer and more detailed understanding of the developmental trajectories of EF components as well as the developmental relations between those components. Finally, attention to sequences can clarify some of the differences among theories of EF development in how conflicting representations are resolved. For example, in Zelazo's CCC theory (Zelazo et al., 2003), two rules are integrated to produce a new overarching rule system. In contrast, in Munakata's (2001) theory, latent representations are maintained, even after active representations strengthen and override latent ones.

Two articles have attempted to extract developmental sequences. Anderson (2002) inferred the following sequence of EF components from an integrative review: attentional control (e.g., inhibition), information processing (e.g., processing speed), cognitive flexibility (e.g., switching), and goal setting. Romine and Reynolds (2005) inferred sequences from the age at which performance leveled off, based on a meta-analysis of ages 5–22 and average effect sizes of age-related change in performance. They found the following sequence: inhibition of perseveration, set maintenance, design fluency, planning, and verbal fluency.

One design for detecting sequences, rarely used in this research area, is to give children slightly different versions of the same task (to try to equate task demands, such as verbal demands and content area), with each task assessing a different component of EF (see Wellman & Liu, 2004, for the successful use of this design for assessing the acquisition sequence for theory-of-mind tasks). If task demands, other than the EF component of interest, are in fact equated, the mean performance on each version or the percent of children who pass each version suggests the order in which the components of EF are acquired. Also, a scalogram analysis could examine how many children pass all of the hypothesized easier versions of the task before a hypothesized more difficult task. That is, if the hypothesized ordering of component tasks from easiest to hardest is A, B, C, D, then the outcome of interest is how many children passed A, B, and C, but not D; A and B but not C and D; and A but not B, C, and D.

Carlson (2005) applied this method in a large sample aged 2–5 to determine the probability of passing common EF tasks at each age. Although the ordering of specific tasks varied from age to age, inhibition tasks routinely were passed earlier than WM tasks and the very hardest tasks consistently involved both inhibition and WM demands (e.g., reverse categorization at age 2, DCCS at age 3, backward digit span at ages 3 and 4, advanced DCCS at ages 5–6).

Finally, longitudinal studies obviously are ideal to detect sequences, but few exist. Several longitudinal studies have examined the TOH or TOL task, with a focus on developmental sequences in strategy use (McNamara et al., 2007), and family, cognitive, school achievement, and social adjustment correlates (Friedman et al., 2007; Jacobson & Pianta, 2007). On inhibitory tasks, one striking sequence identified is that performance on a delay of gratification task at age 4 predicts, and thus may be a developmental precursor for, performance on inhibitory tasks such as the Go-No-Go task at age 18 (Eigsti et al., 2006). In a longitudinal study from ages 2 to 4 (Hughes & Ensor, 2007), EF (an aggregate of inhibition, WM, and shifting tasks) improved with age and showed stable individual differences, indicating the predictive ability of early EF for later EF.

This review suggests several influences on children's level of performance on EF assessments that should be considered when examining sequences. One complication in identifying sequences is that estimates of performance, and thus developmental trajectories, may vary, depending on what aspect of performance is scored and how it is scored (e.g., Baker et al., 2001; Huizinga et al., 2006). Moreover, the apparent developmental order of two aspects (A then B) of EF actually could be due to the greater performance demands of B. Reducing the demands of B would reverse the sequence. For this reason the design suggested above, with different versions of the same task, ideally would use similar levels of task difficulty. These measurement issues obviously become even more challenging when assessing the sequence of a task requiring one EF component and a task requiring two components.

Mechanisms identified. The studies reviewed here suggest several likely developmental mechanisms (biological and environmental) of EF development for the focus of future research. Consistent with most developmental EF theories, the assessment of brain activity is important because it provides clues about how developing EFs become organized, as seen in changes in neural networks and increased localization. Such research is well underway, although this research focuses on inhibition and WM, and rarely shifting. One influential developmental model (Posner & Rothbart, 2007) proposes that the development of the anterior attention system—the executive attention network—during preschool is important for regulating other brain networks.

Large neural and behavioral changes in the components during the preschool years are followed by more gradual, fine-grained improvements later. However, establishing *specific* links between brain changes and changes in behavior has proven to be more difficult. Also, distinguishing individual performance differences in brain activity from maturational differences continues to challenge neuroimaging researchers (Thomas & Tseng, 2008). One promising approach is to include several brain and behavioral measures and examine which ones change from one age to another and which do not. For example, LaVallee, Muenke, Robertson, and Watamura (2007) compared EEG responses, reaction time, number of gaze shifts, and accuracy on a modified Stroop task (e.g., see a boy, hear the word "girl") in 3- versus 4-year-olds. Another promising approach is one (Morton et al., 2009) in which age differences in neural correlates of component processes (shifting vs. processing of conflict) were examined separately.

Two points should be made about brain-based changes as possible mechanisms. First, any change in brain function could be either a cause (i.e., neural maturation provides a mechanism of development) or an effect (i.e., EF behaviors lead to brain changes) or both. Second, cognitive neuroscience work would contribute more to the existing behavioral literature if it were tied to theoretical developmental issues such as quantitative versus qualitative change, degree of generalization of new EF skills, and domain-specific versus domain-general EF skills.

As for experiential-based mechanisms of EF development, several studies reviewed (e.g., Crone, 2007; Crone et al., 2006; Davidson et al., 2006; Hogan et al., 2005; Somsen, 2007) suggest that metacognition may play an important role during the school-age years and adolescence. Examples are an awareness of inhibition failures and the subsequent adjustments in response to avoid future errors, as well as slowing one's response in order to ensure high accuracy on a switching task. Current developmental EF theories, because they are based mainly on development in the first 5 years of life, are limited in this respect. Post-preschool detection of success or failure of one's current rule (Zelazo), latent representation (Munakata), mental set or prepotent behavior (Diamond), or focus of attention (Posner & Rothbart) may be important metacognitive developments that would extend these theories based on preschoolers to older children. The relevant aspects of metacognition may vary from one component to another. For example, it seems likely that knowing to slow down is particularly important in theories that emphasize inhibition, whereas detecting errors and considering alternative responses may be particularly important in theories emphasizing shifting.

Several other suspected contributors to EF include practice (e.g., McNamara et al., 2007), intense motor activity (Bell et al., 2007; Campbell, Eaton, & McKeen, 2002; Davis et al., 2007), language (Bell et al., 2007; for preschoolers only; Kray, Eber, & Lindenberger, 2004; Wolfe & Bell, 2004), bilingualism (Carlson & Meltzoff, 2008), maternal education and parenting (Friedman et al., 2007), and theory of mind (Hughes & Ensor, 2007; Perner & Lang, 2000)—understanding that mental states exist and affect behavior. Moreover, cultural differences, such as the earlier acquisition of EF in Chinese than U.S. children (e.g., Sabbagh et al., 2006), suggest that cultural values, perhaps as expressed in practices at school, may affect the development of EF.

Conclusions

Unlike previous reviews focused on preschoolers' EF, this review focused on EF across a much larger age span. This perspective permitted an examination of EF in light of central developmental issues such as the form of developmental trajectories, sequences of acquisition within EF development, qualitative and quantitative change, and developmental mechanisms at both behavioral and neural levels. Based on this framework of developmental issues, the key components of needed future research include: (a) use of a wide age range and comparable tasks to reveal the form of developmental trajectories of each EF component, (b) examination of several EF components so that relations among components can be examined, and (c) assessment of possible mechanisms of development. Such designs would move EF research from its current state—strong theoretical and empirical work on preschoolers, scattered nonintegrated work on older children, and emphasis on description of age differences—to a truly developmental account. This account would provide a developmental theoretical focus to cognitive neuroscience studies of children's EF. In turn, a more theoretically based developmental cognitive neuroscience would provide greater constraints on developmental issues and theories of EF than we have now.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71–82.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia, 44*, 2259–2269.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baker, K., Segalowitz, S. J., & Ferlisi, M.-C. (2001). The effect of differing scoring methods for the Tower of London task on developmental patterns of performance. *The Clinical Neuropsychologist, 15*, 309–313.

Barkley, R. A., Edwards, G., Laneri, M., Fletcher, K., & Metevia, L. (2001). Executive functioning, temporal discounting, and sense of time in adolescents with attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD). *Journal of Abnormal Child Psychology, 29*, 541–556.

Bell, M. A., Wolfe, C. D., & Adkins, D. R. (2007). Frontal lobe development during infancy and childhood. In D. J. Coch, K. W. Fischer, & G. Dawson (Eds.), *Human behavior learning, and the developing brain: Typical development* (pp. 247–276). New York: Guilford.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem solving task: Enhancing clinical and research implementation. *Journal of Experimental and Clinical Neuropsychology, 25*, 586–604.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180–200.

Beveridge, M., Jarrold, C., & Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant and Child Development, 11*, 107–123.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*, 571–593.

Campbell, D. W., Eaton, W. O., & McKeen, N. A. (2002). Motor activity level and behavioral control in young children. *International Journal of Behavioral Development, 26*, 289–296.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–616.

Carlson, S. M., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282–298.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Casey, B. J., Amso, D., & Davidson, M. C. (2006). Learning about learning and development with modern imaging technology. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (Vol. 21, pp. 513–533). Oxford, UK: Oxford University Press.

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., et al. (1997). A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience, 9*, 835–847.

Cepeda, N. J., Kramer, A. F., & Gonzales de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology, 37*, 715–730.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology, 31*, 103–128.

Cragg, L., & Nation, K. (2008). Go or no-go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science, 11*, 819–827.

Crone, E. A. (2007, April). *Development of feedback processing and performance monitoring.* Paper presented at the meeting of the Society for Research in Child Development, Boston.

Crone, E. A., Somsen, R. J. M., Zanolie, K., & Van der Molen, M. W. (2006). A heart rate analysis of developmental change in feedback processing and rule shifting from childhood to early adulthood. *Journal of Experimental Child Psychology, 95*, 99–116.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

Davis, C. L., Tomporowski, P. D., Boyle, C. A., Waller, J. L., Miller, P. H., Naglieri, J., et al. (2007). Effects of aerobic exercise on overweight children's cognitive functioning: A randomized controlled trial. *Research Quarterly for Exercise & Sport, 78*, 510–519.

Davis, C. L., Tomprowski, P. D., McDowell, J. E., Austin, B. P., Miller, P. H., Yanasak, N. E., et al. (in press). Exercise improves executive function and alters neural activation in overweight children: A randomized controlled trial. *Health Psychology*.

DeMarie-Dreblow, D., & Miller, P. H. (1988). The development of children's strategies for selective attention: Evidence for a transition period. *Child Development, 59*, 1504–1513.

D'Esposito, M., & Postle, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia, 37*, 1303–1315.

Diamond, A. (1990a). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*, 637–676.

Diamond, A. (1990b). The development and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences, 608*, 267–317.

Diamond, A. (2006). The early development of executive functions. In E. Bialystock & F. I. M. Craik (Eds.), *The early development of executive functions. Lifespan cognition: Mechanisms of change* (pp. 70–95). Oxford, UK: Oxford University Press.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to ''do as I say, not as I do.'' *Developmental Psychobiology, 29*, 315–334.

Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., et al. (2006). A shift from diffuse to focal cortical activity with development. *Developmental Science, 9*, 1–20.

Eigsti, I., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M., et al. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science, 17*, 478–484.

Fischer, B., Biscaldi, M., & Gezeck, S. (1997). On the development of voluntary and reflexive components in human saccade generation. *Brain Research, 754*, 285–297.

Flavell, J. H. (1972). An analysis of cognitive-developmental sequences. *Genetic Psychology Monographs, 82*, 279–350.

Friedman, S. L., Vandergrift, N., Spieker, S., Hirsh-Pasek, K. A., Keating, D. P., Park, Y., et al. (2007, April). *First through fifth grade trajectories of Tower of Hanoi planning: family and cognitive predictors; math and reading correlates.* Poster presented at the meeting of the Society for Research in Child Development, Boston.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½-7 years old on a Stroop-like day-night test. *Cognition, 53*, 129–153.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 8174–8179.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development, 58*, 601–622.

Hogan, A. M., Vargha-Khadem, F., Kirkham, F. J., & Baldeweg, T. (2005). Maturation of action monitoring from adolescence to adulthood: an ERP study. *Developmental Science, 8*, 525–534.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*, 1447–1459.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017–2036.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7*, 942–951.

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology, 26*, 403–422.

Jacobson, L. A., & Pianta, R. C. (2007, April). *Executive function skills and children's academic and social adjustment to sixth grade.* Poster presented at the biennial meeting of the Society for Research in Child Development, Boston.

Johnstone, S. J., Dimoska, A., Smith, J. L., Barry, R. J., Pleffer, C. B., Chiswick, D., et al. (2007). The development of stop-signal and go/nogo response inhibition in children aged 7-12 years: Performance and event-related potential indices. *International Journal of Psychophysiology, 63*, 25–38.

Jonkman, L. M. (2006). The development of preparation, conflict monitoring and inhibition from early childhood to young adulthood: A go/nogo ERP study. *Brain Research, 1097*, 181–193.

Jonkman, L. M., Lansbergen, M., & Stauder, J. E. A. (2003). Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. *Psychophysiology, 40*, 752–761.

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*, 407–428.

Klimkeit, E. I., Mattingley, J. B., Sheppard, D. M., Farrow, M., & Bradshaw, J. L. (2004). Examining the development of attention and executive functions in children with a novel paradigm. *Child Neuropsychology, 10*, 201–211.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience, 14*, 1–10.

Kray, J., Eber, E., & Lindenberger, U. (2004). Age differences in executive functioning across the lifespan: The role of verbalization in task preparation. *Acta Psychologica, 115*, 143–165.

Kwon, H., Reiss, A. L., & Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proceedings of the National Academy of Sciences USA, 99*, 13336–13341.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia, 44,* 2139–2148.

LaVallee, S. E., Muenke, H. S., Robertson, S. S., & Watamura, S. E. (2007, March). *The development of executive control: Behavioral and neural activity in a proposed pediatric stroop task.* Poster presented at the meeting of the Society for Research in Child Development, Boston.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21,* 59–80.

Lehto, J. E., & Uusitalo, A.-K. (2006). Rule detection in preschool-aged children. *European Journal of Developmental Psychology,* 3, 209–221.

Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience, 114,* 1291–1311.

Liston, C., Watts, R., Tottenham, N., Davidson, M. C., Niogi, S., Ulug, A. M., et al. (2006). Frontostriatal microstructure modulates efficient recruitment of cognitive control. *Cerebral Cortex, 16,* 553–560.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development, 76,* 697–712.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36,* 273–293.

McNamara, J. P. H., DeLucca, T. L., & Berg, W. K. (2007, March). *Children's executive functioning: A longitudinal and microgenetic study.* Poster presented at the meeting of the Society for Research in Child Development, Boston.

Miyake, A. (2009, May). *Individual differences in executive function: Basic findings and implications for self-regulation research.* Paper presented at the meeting of the Association for Psychological Science, San Francisco.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Morton, J. B., Bosma, R., & Ansari, D. (2009). Age-related changes in brain activation associated with dimensional shifts of attention: an fMRI study. *NeuroImage, 46,* 336–358.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5,* 309–315.

Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research, 121,* 391–400.

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of Cognitive Neuroscience, 16,* 1227–1233.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126,* 220–246..

O'Hare, E. D., & Sowell, E. R. (2008). Imaging developmental changes in gray and white matter in the human brain. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 23–38). Cambridge, MA: MIT Press.

Olson, E. A., & Luciana, M. (2008). The development of prefrontal cortex functions in adolescence: Theoretical models and a possible dissociation of dorsal versus ventral subregions. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 575–590). Cambridge, MA: MIT Press.

Perner, J., & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 150–181). New York: Oxford University Press.

Polich, J. (1987). Task difficulty, probability and interstimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 63,* 251–259.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology, 58,* 1–23.

Rennie, D. A. C., Bull, R., & Diamond, A. (2004). Executive functioning in preschoolers: Reducing the inhibitory demands of the Dimensional Change Card Sort task. *Developmental Neuropsychology, 26,* 423–443.

Roberts, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology, 12,* 105–126.

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe function: Findings from a meta-analysis. *Applied Neuropsychology, 12,* 190–201.

Rubia, K., Smith, A. B., Woolley, J., Nosarti, C., Heyman, I., Taylor, E., et al. (2006). Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. *Human Brain Mapping, 27,* 973–993.

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 14931–14936.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81.

Scherf, K. S., Sweeney, J. A., & Luna, B. (2006). Brain basis of developmental change in visuospatial working memory. *Journal of Cognitive Neuroscience, 18,* 1045–1058.

Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology, 26*, 445–464.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*, 207–218.

Simpson, A., & Riggs, K. J. (2005). Inhibitory and working memory demands of the day-night task in children. *British Journal of Developmental Psychology, 23*, 471–486.

Somsen, R. J. M. (2007). The development of attention regulation in the Wisconsin Card Sorting Task. *Developmental Science, 10*, 664–680.

Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and behavioral science. *Current Directions in Psychological Science, 16*, 55–59.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: Evidence for a dual systems model. *Developmental Psychology, 44*, 1764–1778.

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of frontal lobes. *Psychological Bulletin, 95*, 3–28.

Thomas, K. M., & Tseng, A. (2008). Functional MRI methods in developmental cognitive neuroscience. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 311–324). Cambridge, MA: MIT Press.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition, 41*, 231–242.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology, 35*, 205–213.

Wolfe, C. D., & Bell, M. A. (2004). Working memory and inhibitory control in early childhood: Contributions from physiology, temperament, and language. *Developmental Psychobiology, 44*, 68–83.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(Serial No. 274).

Child Development, November/December 2010, Volume 81, Number 6, Pages 1661–1669

Young Children Selectively Avoid Helping People With Harmful Intentions

Amrisha Vaish, Malinda Carpenter, and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology

Two studies investigated whether young children are selectively prosocial toward others, based on the others' moral behaviors. In Study 1 ($N = 54$), 3-year-olds watched 1 adult (the actor) harming or helping another adult. Children subsequently helped the harmful actor less often than a third (previously neutral) adult, but helped the helpful and neutral adults equally often. In Study 2 ($N = 36$), 3-year-olds helped an actor who intended but failed to harm another adult less often than a neutral adult, but helped an accidentally harmful and a neutral adult equally often. Children's prosocial behavior was thus mediated by the intentions behind the actor's moral behavior, irrespective of outcome. Children thus selectively avoid helping those who cause—or even intend to cause—others harm.

Humans are not equally prosocial toward all individuals. Instead, we select the beneficiaries of our prosocial acts from among multiple potential beneficiaries. One crucial criterion for this selection is the beneficiary's helpful or harmful behavior: We help and affiliate with helpful individuals and shun or punish harmful individuals (e.g., Fehr & Gächter, 2002; Krebs, 2008). To date, almost no work has examined whether children make such selections. We do know that infants' social preferences for nonhuman agents are determined by those agents' helping or hindering behaviors (Hamlin, Wynn, & Bloom, 2007) and that children begin helping others early in life (Warneken & Tomasello, 2006, 2007), but relations between the two have not yet been assessed. Later, in the preschool years, children's abstract moral reasoning about story characters is related to their everyday prosocial behavior (Eisenberg-Berg & Hand, 1979; Janssens & Dekoviç, 1997). In one particularly relevant study, Olson and Spelke (2008) asked 3.5-year-olds to help a protagonist doll decide how to allocate resources to other dolls. Children allocated more of the protagonist doll's resources to a doll who was generous to the protagonist doll (direct reciprocity) or generous to a different doll (indirect reciprocity) than to a nongenerous doll, thus demonstrating some appreciation of direct and indirect reciprocity.

Olson and Spelke's (2008) study could also be interpreted as showing that young children consider others' moral behaviors when deciding whom to share with. However, sophisticated moral judgments take into account several factors, including not only the consequences of but also the intentions behind a behavior (see Cushman, 2008; Piaget, 1932/1997; Turiel, 2006; Turiel, Killen, & Helwig, 1987; Weiner & Peter, 1973). Thus, to show that children's prosocial behavior is mediated by others' moral behavior, one needs to show that the prosocial behavior is mediated not only by the outcomes of but also the intentions behind others' actions. In the present work, we assessed whether young children show differential prosocial behavior toward agents who vary in the helpful or harmful intentions underlying their behavior.

A secondary issue, also raised by Olson and Spelke (2008), is that children's responses on behalf of others (the protagonist doll in their study) may not correspond to children's responses when they themselves must decide whom to share with or help. Our first study thus tested whether children respond differentially to others' harmful or helpful behaviors even when they themselves are the ones sharing or helping. The primary question of the second study was whether intentionality mediates

---

The authors thank the children and parents who participated in the study. They also thank Eva Leermann and Elvira Plath for help with data collection, Sandy Kennert and Jutta Schulze for help with coding, and Tobias Grossmann for helpful discussions and comments on the manuscript. This research was partially supported by a grant from the Excellence Cluster 302: Languages of Emotion.

Correspondence concerning this article should be addressed to Amrisha Vaish, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. Electronic mail may be sent to vaish@eva.mpg.de.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0003

children's judgments of others as worthy recipients of prosocial acts. Note that in both studies, children witnessed but were not the recipients of the harmful or helpful behaviors. Using such third-party interactions circumvents ethical issues and ensures that children's subsequent actions are responses to the actor's behavior rather than a result of their own emotional responses to being helped or harmed (see Olson & Spelke, 2008).

The general procedure of both studies was as follows: In a between-subjects design, after warming up with three adults, the children viewed four familiarization trials during which one adult (the actor) behaved in a condition-specific way (e.g., harmfully or helpfully) toward another adult (the recipient). The four familiarization trials were followed by one helping test in which children were presented with a forced choice between helping the actor or the third (familiar but neutral) adult. The dependent measure in both studies was thus instrumental helping, which is relatively easy to elicit without explicit verbal instructions (Warneken & Tomasello, 2006, 2007). A forced choice was included because young children are highly motivated to help (see Warneken & Tomasello, 2009) and might help a harmful actor if that were their only option; the forced-choice paradigm gave them the option of not helping the actor. Finally, once children had helped one person in the helping test, they were given the opportunity to help the other person if they chose so.

Study 1

In Study 1, children saw the actor harming, helping, or, as a baseline measure, behaving neutrally toward the recipient. We predicted that children's subsequent helping would be mediated by the actor's behaviors toward the recipient.

Method

Participants

Participants were 3-year-olds ($N = 54$; 9 girls and 9 boys per condition) between 35 months 21 days and 38 months 17 days ($M = 37$ months 11 days, $SD = 22.46$ days) from a medium-sized German city. Fifteen additional children were tested but excluded because of indecisiveness or unclear responses during the helping test (6; see the Procedure section for a description of a clear helping response), fussiness or inattentiveness (4), parents not following instructions (4), and experimenter error (1). Children were recruited from a database of parents who volunteered to participate in child development studies. Participants were all native German speakers and came from mostly middle-class backgrounds. Although precise information about the sample's ethnicity is not available, about 98% of the population from which the sample was drawn is native German. All children were seen in a child laboratory for a 45-min ''play'' session. The same three adult female experimenters played the same roles (of recipient, actor, and neutral person) for all children.

Materials and Setting

The setup is shown in Figure 1. Each child saw four familiarization trials during which the child, recipient, and actor sat around a table; the neutral person sat to the side, visible to the child, reading a magazine. The following materials were used during familiarization trials: two necklaces with colorful beads, two belts with colorful beads, blank sheets of paper and a color pencil, and a blue and a red ball of clay in a container.

In the subsequent helping test, two identical color-matching games were used which consisted of a box with four holes marked by different colors into which balls of matching colors could be placed. The child stood at a predetermined location and the color-matching games' boxes were placed 2 m from and on either side of the child (see Figure 1). Next to each box lay three of the four balls required for each game; a blue ball was missing from both sides. One blue ball lay in the middle, 1 m from the child. The parents sat directly behind the child, the actor and neutral person sat next to their game (side counterbalanced across children), and the recipient stood to the side of and behind the child, facing away from the interaction, and kept time.

Procedure

After warming up with the children, the recipient introduced the actor and neutral person to children as her friends. The actor and neutral person then warmed up with children, being careful to interact equally and similarly with them. Throughout, the recipient wore one of the necklaces and belts described before. After about 10 min in the warm-up room, the recipient told children that she would like to show them the toys in a different room, and escorted the children and parents to the testing room, followed by the actor and the neutral person.

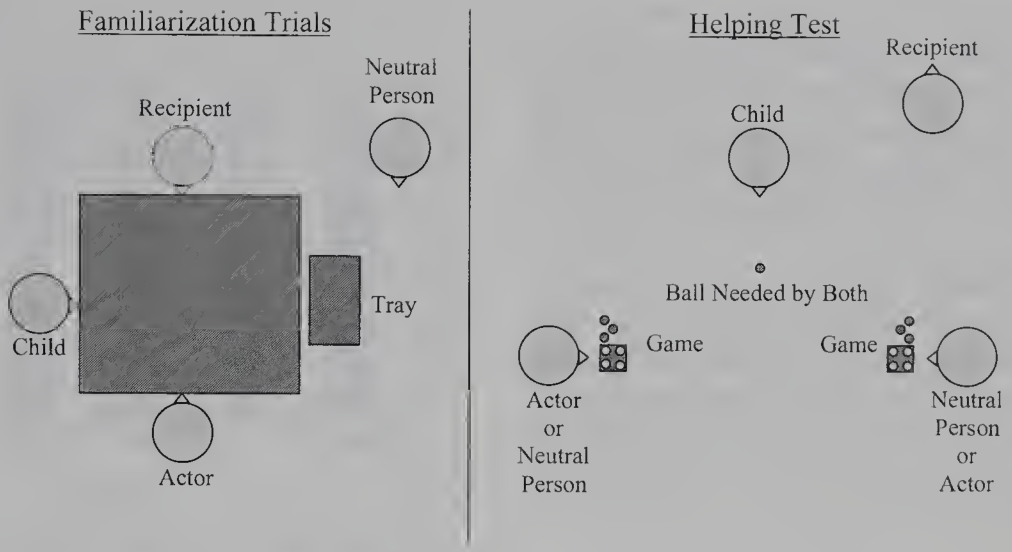


Figure 1. Experimental setup for all conditions.

The study was between-subjects with three conditions (harm, help, and baseline). All conditions had four familiarization trials followed by one helping test. Each familiarization trial (adapted from Vaish, Carpenter, & Tomasello, 2009) began with a 45-s presentation in which the recipient presented one of the following objects (order counterbalanced):

Necklace and belt. The recipient admired and showed off her necklace or belt. A second necklace and belt lay on a tray, visible but inaccessible to children.

Picture. On one sheet of paper from the stack, the recipient drew a picture, proudly commenting on how pretty it was and how happy it made her.

Clay. Using one ball of clay, the recipient made a bird, commenting as in the picture trial. The second clay ball remained in the container.

These 45-s presentations ended differently depending on the condition: In the harm condition, the recipient placed her possession on the table and the actor said in a mildly aggressive tone, "I'm going to take [or tear or break] this now," and put on the necklace or belt, or tore up the picture or bird and threw the bits into a bin. The actor did not display aggression in her facial expression (which was neutral while she spoke and during her actions), nor in any other way before or during her actions. The recipient watched the actor sadly but silently. After 15 s, the recipient assumed a neutral expression and began the next demonstration.

In the help condition, in contrast, the recipient was accident-prone: Instead of placing her possession on the table, she accidentally dropped her necklace, detached beads from her belt, tore her picture, or damaged her clay bird, and was sad about each mishap. The actor said sympathetically, "I'll get [or fix] it," then retrieved or repaired the object. The recipient watched the actor sadly but silently (as in the harm condition). After 15 s, the actor placed the object on the table and smiled, and the recipient happily took it, put it aside, and neutrally began the next demonstration.

The baseline condition established the appropriate chance level for analyses of the other conditions and thus involved the actor behaving neutrally. After necklace and belt demonstrations, the recipient told the children that the necklace (or belt) on the tray was not hers. The actor then commented on the necklace or belt on the tray ("The necklace has so many beads" or "This belt can be adjusted here") and counted the beads or adjusted the belt. After picture and clay demonstrations, the actor said, "These [remaining] sheets are all blank" or "This [second ball of] clay is stuck to the container," and proceeded to examine the paper or unstick the clay. The actor's comments and expression were neutral and the recipient watched the actor neutrally. After 15 s, the actor put the object away and the recipient began the next demonstration.

In each condition, the familiarization trials were followed by a helping test in which the actor and neutral person simultaneously but individually played their respective color-matching game. After placing the three available balls in their slots, both simultaneously reached for the ball in the middle and maintained this reach, looking only at the ball. If children did not act within 15 s, the recipient silently cued parents to ask children to "give the ball." If this was ineffective, parents asked children to point to the person they wanted to give the ball to. As a last resort, the recipient sat between the actor and neutral person, held up the ball, and asked children to give it or point to the person they wanted to give it to. Whomever children chose gratefully accepted the ball. Only placing the ball in or near one person's hand or pointing to one person were considered clear responses. If children did not respond within 1 min, the test was ended.

After the helping test, children received a second ball to hand to whomever they chose. This helped resolve the situation by giving children the chance to help both individuals. Finally, in the harm condition, the actor made amends (e.g., by returning the recipient's necklace and belt) and apologized, and the recipient accepted the apology.

Coding and Reliability

During testing, the primary author coded to whom children chose to give the ball. A second coder (blind to condition and hypotheses) coded this measure from videotapes for a random 50% of children (9 per condition). Agreement was perfect, $\kappa = 1.0$. For the same 50%, before giving the code, the second coder judged (using relevant cues such as one person reaching slightly further than the other) whether either person was likelier to receive the ball. She judged the actor as likelier in seven cases and the neutral person in five cases, but these numbers were unrelated to condition and to who received the ball (Fisher's exact tests, *p*s = .47 and 1.00, respectively).

Results

As preliminary analyses revealed no gender effects, gender was not included in further analyses. All reported *p* values are two-tailed. In the baseline condition, 12 of 18 children (67%) helped the actor; 67% was thus the appropriate test proportion for the other conditions. Following Hamlin et al. (2007), we used binomial tests to analyze the experimental conditions. In the harm condition, a significantly lower proportion of children than 67% helped the actor (4 of 18, or 22%; binomial probability, $p < .0005$). This difference did not emerge in the help condition (11 of 18 children, or 61%, helped the actor; binomial probability, $p = .760$; see the first three bars of Figure 2).

Secondary analyses using chi-square tests revealed that the harm condition differed significantly from the baseline condition, $\chi^2(1, N = 36) = 7.20$, $p = .007$, and from the help condition, $\chi^2(1, N = 36) = 5.60$, $p = .018$. The help and baseline conditions did not differ, $p = .729$.

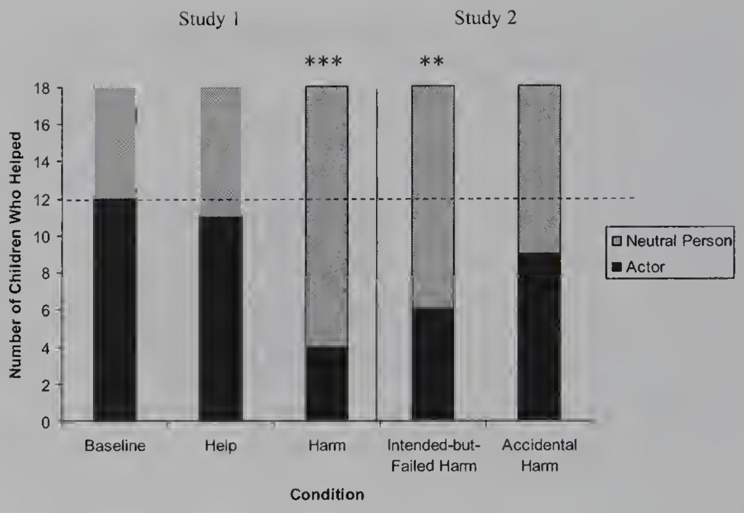


Figure 2. Number of children who helped the actor or the neutral person in Studies 1 and 2.
Note. The dashed line represents chance (67%) as established by the baseline condition.
$^{**}p < .01$. $^{***}p < .0005$.

An alternative explanation for the results of the harm condition is that children refrained from helping the actor because they were afraid of her (e.g., because of her mildly aggressive tone of voice). However, the phase in which children received a second ball suggests otherwise: All 14 children who had first helped the neutral person gave the second ball to the actor, indicating they were not afraid of her but, when forced to choose during the helping test, chose not to help her.

Discussion

This study showed that 3-year-olds take into account others' harmful behaviors toward third parties when deciding whether or not to help them. This extends Olson and Spelke's (2008) findings to children's own prosocial behavior and suggests that children select whom to help and selectively direct resources away from harmful people.

The actor's harmful but not her helpful actions impacted children's behavior. This suggests a negativity bias, that is, a greater impact of negative than of positive information. This bias has been extensively documented in adults (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) and in children's social-emotional development (Vaish, Grossmann, & Woodward, 2008). Our findings suggest a similar bias in children's moral development: Children's prosocial behavior is decreased toward a harmful individual but not increased toward a helpful individual. This result is consistent with the research on children's moral judgments that indicates that children correctly identify "bad" acts substantially earlier than "good" acts (Hill & Hill, 1977; Rhine, Hill, & Wandruff, 1967; see Karniol, 1978; Smetana, 2006; Turiel, 2006, for discussions). Future studies should further explore the origins and implications of the negativity bias in moral development (see Aloise, 1993; Leslie, Knobe, & Cohen, 2006, for relevant work).

Notably, although most children in the harm condition helped the neutral person first, they handed a second ball to the actor. This concords with Olson and Spelke's (2008) finding that although 3.5-year-olds allocated resources selectively when resources were limited, when resources were sufficient, children divided them equally among recipients, displaying a sense of fairness. When a second ball was available, children in our study also displayed fairness, or at the very least, they were willing to help the actor when she was the only one who needed help.

In Study 1, young children helped a harmful individual less, but whether this was in response to the individual's harmful intentions or to the negative outcomes she caused remains unclear because the harm condition featured both intentional harm and negative consequences. The critical question of why children helped a harmful actor less was addressed in Study 2.

Study 2

The intention–outcome distinction is essential to the study of moral development, given that sophisticated moral judgments are thought to rely on not only the consequences of but also the intentions behind others' behavior (Cushman, 2008; Karniol, 1978; Piaget, 1932/1997; Turiel et al., 1987). Developmental work suggests that when intentions and outcomes are pitted against each other, children around 5 years of age and above reliably use the perpetrator's intentions when making moral judgments and assigning punishment; prior to this, they rely largely on outcome information (e.g., Miller & McCann, 1979; Wellman, Larkey, & Somerville, 1979; Zelazo, Helwig, & Lau, 1996). Much of this past work has assessed children's verbal evaluations of hypothetical moral transgressions, which may not be optimal for very young children whose language skills are limited and which do not always correspond to children's actual behavior (Astington, 2004; Darley & Shultz, 1990; Wainryb, Brehl, & Matwin, 2005). We asked in our second study whether young children's ability to respond to the intentions behind moral transgressions might be evident in their prosocial behavior.

In Study 2, we teased apart intentions and outcomes using two new conditions: In the intended-but-failed harm condition, the actor intended but was unable to harm the recipient (non-negative outcome, negative intention), and in the accidental harm condition, the actor accidentally harmed the recipient (negative outcome, non-negative intention). If the intentions behind others' harmful actions mediate children's prosocial behavior, children should subsequently help the actor less than the neutral person in the intended-but-failed harm condition but not in the accidental harm condition.

Method

Participants

Participants were 3-year-olds ($N = 36$; 9 girls and 9 boys per condition) between 35 months 18 days and 38 months 15 days ($M = 37$ months 7 days, $SD = 24.93$ days). Seven additional children were tested but excluded because of indecisiveness or unclear responses during the helping test (3), fussiness or inattentiveness (2), parent not following instructions (1), and experimenter error (1). Participant recruitment, the sample's ethnicity and socioeconomic backgrounds, and the duration of sessions were the same as in Study 1. The same experimenters played the same roles as in Study 1.

Materials and Setting

The materials were the same as in Study 1 with minor changes. In the intended-but-failed harm condition, the sheets of paper had a border of transparent tape such that they could not easily be torn. In the accidental harm condition, a set of beads was wrapped around the belt such that it could easily fall off, and the sheets of paper had an inconspicuous rip such that they could easily tear. The setting was identical to Study 1.

Procedure

The warm-up session was identical to Study 1. The two between-subjects conditions of Study 2 again involved four familiarization trials and a helping test. The recipient's presentations during familiarization trials were similar to Study 1 except that she decorated her clay bird with beads. The four 45-s presentations ended with the recipient placing her possession on the table. In the intended-but-failed harm condition (like the harm condition), the actor then said mildly aggressively, "I'm going to tear [or break] this now," and attempted (with obvious effort) to tear up the picture, dislodge the beads from the clay bird, or break the necklace or belt. She did not display aggression facially or in any other way before or during her actions. Importantly, she was unable to cause harm as she could not tear through the tape around the paper, the beads were lodged too deeply into the clay, and pulling on the beads of the necklace and belt did not break them off. The recipient watched the actor sadly but silently. After 15 s, the actor gave up and placed the object back on the table. The recipient examined the object, smiled briefly (to indicate satisfaction that it was intact), placed it to the side, and neutrally began the next presentation.

In the accidental harm condition, the actor accidentally destroyed the recipient's objects. Thus, the actor admired the picture but accidentally tore it while returning it to the recipient. Similarly, the beads fell off the belt and the knot of the necklace came undone while she was admiring these objects. In the clay situation, the actor was looking at her watch when the bird was placed on the table so that when she turned to admire the bird, her arm collided with it and broke it. After each mishap, the actor said in a sorry tone of voice, "I didn't want that to happen" or "That wasn't on purpose" (in alternating order, beginning with the former). The actor apprehensively examined the broken object but did not repair it or apologize, and the recipient watched the actor silently but sadly. After 15 s, the actor placed the object to the side and the recipient neutrally began the next presentation.

As in Study 1, familiarization trials were followed by a helping test and a final phase in which children could hand over a second ball. Finally, in the intended-but-failed harm condition, the actor apologized and the recipient accepted the apology.

Coding and Reliability

Coding and reliability were conducted as in Study 1. Agreement was perfect between the primary and second coders' coding of whom children gave the ball to, $\kappa = 1.0$. The second coder also judged that the actor was more likely to receive the ball in one case and the neutral person in four cases, but these numbers were unrelated to condition and to who received the ball (Fisher's exact test, *p*s = .40 and 1.00, respectively).

Results

As preliminary analyses revealed no gender effects, gender was not included in further analyses. All reported p values are two-tailed. We again used the test proportion (67%) established by the baseline condition in Study 1. Binomial tests revealed that in the intended-but-failed harm condition, a significantly lower proportion of children than 67% helped the actor (6 of 18, or 33%; binomial probability, $p = .008$). This difference did not emerge in the accidental harm condition (9 of 18 helped the actor, $p = .204$; see the last two bars of Figure 2).

Chi-square analyses revealed that the intended-but-failed harm condition differed significantly from the baseline condition of Study 1, $\chi^2(1, N = 36) = 4.00, p = .046$, but not from the harm condition, $p = .457$. In contrast, the accidental harm condition differed marginally from the harm condition, $\chi^2(1, N = 36) = 3.01, p = .083$, but not from the baseline condition, $p = .310$.

Results from the final (second ball) phase again suggest that children were not afraid of the actor in the intended-but-failed harm condition: Of the 12 children who helped the neutral person during the helping test, 10 handed the second ball to the actor.

Discussion

In Study 2, 3-year-olds decreased their prosocial behavior toward a person who had harmful intentions toward a third party even if she did not cause negative outcomes. Moreover, children did not significantly decrease their prosocial behavior toward a person who caused negative outcomes without harmful intentions. These findings are the first evidence that by age 3, children selectively reduce prosocial behavior toward intentionally harmful—and thus morally blameworthy—individuals regardless of the consequences of those individuals' actions.

Study 2 contributes to the moral development literature as it suggests that as early as 3 years, children respond differently to intentionally versus unintentionally caused harm. In prior work in which intentions and outcomes were pitted against each other in hypothetical moral transgressions, children reliably used a perpetrator's intentions in their moral evaluations starting around 5 years of age (e.g., Wellman et al., 1979; Zelazo et al., 1996). Interestingly, however, Imamoğlu (1975) found that even when older children (5-year-olds) fail to differentially evaluate intentional versus accidental acts, they respond differentially on other measures such as like–dislike judgments of the perpetrator. Thus, young children's understanding of a perpetrator's intentions may be more apparent in their liking of or willingness to help the perpetrator than in their verbal evaluations of the transgression. Affiliation and helping may thus be precursors to and important facets of children's moral evaluations. More generally, nonverbal, nonhypothetical behavioral measures complement verbal measures of children's judgments and reasoning about intentions. Future work should accordingly employ both verbal and behavioral measures focusing on both the act and the actor to obtain a fuller picture of children's moral judgment-making.

General Discussion

The present studies demonstrate that young children's prosocial behavior is mediated by others' moral behavior. In Study 1, 3-year-olds helped a harmful adult less than a neutral adult, extending Olson and Spelke's (2008) findings to children's own prosocial behavior. The important new finding from Study 2 was that 3-year-olds grasped the intentions behind harmful behavior and selectively decreased their prosocial behavior toward the actor if and only if she could be held morally responsible for her actions (i.e., when she had harmful intentions), even if she was unsuccessful in causing harm. Thus, by age 3, children selectively withhold help from morally blameworthy individuals.

It could be argued that children in the harm and intended-but-failed harm conditions helped the actor less because of her mildly aggressive tone of voice rather than her harmful behavior. However, some aggression is likely a reliable cue to intentional as opposed to accidental harm. Indeed, our rationale for including mild aggression was that having the actor speak entirely neutrally would create an unnatural harming situation in which the actor's intentions would be ambiguous. For instance, children might infer from a neutral voice that the actor had not registered or did not remember that the objects belonged to the recipient and she thus did not intend harm. Thus, although future studies could control for the actor's aggression, we argue that some mild aggression is naturally linked with intentionally harmful behavior and retaining it creates a more ecologically valid situation.

There are also several reasons to believe that the actor's mild aggression alone does not explain the results. First, the actor was only aggressive in her tone of voice and only before the transgressions, not in any other way or at other times during the procedure. Also, nearly all children in both the harm and intended-but-failed harm conditions gave a second ball to the actor, suggesting that her aggression had not made them afraid of her (though it is possible that children were afraid of the actor but their fear did not stop them from helping her in this context). Finally, in a recent study exploring children's protest against moral transgressions (Vaish, Missana, & Tomasello, in press), a puppet proclaimed her harmful intentions (to tear another puppet's drawing because she did not like it) in the same neutral, nonaggressive way in which she proclaimed her nonharmful intentions (to tear a blank piece of paper because she did not like it). Three-year-olds nevertheless protested more in the harmful than in the neutral case. Thus, young children do recognize moral transgressions even in the absence of aggression, at least when the transgressor provides another reason for the transgression (not liking the drawing). It is thus unlikely that our findings are due solely to the slightly aggressive way in which the actor spoke before her harmful actions.

In the present studies, children directed less helping toward a person who harmed or intended to harm a third party. A similar result emerges in adults: In economic games, adults punish individuals who show unfair behavior even when they were not themselves affected by that behavior (Fehr & Fischbacher, 2003, 2004; Singer et al., 2006). Our findings show that the ability to identify harmful individuals and to withhold help from them emerges early in childhood, although it remains unclear whether children actively punished the harmful actor or more passively shunned or avoided her.

Our results are limited to children's interactions with adults and may not generalize to children's interactions with their peers (see, e.g., Killen, 1991). Numerous authors have observed that child–child interactions provide a rich and unique context within which children develop a sense of fairness, equality, and justice (e.g., Arsenio & Lover, 1995; Damon & Killen, 1982; Dunn, Cutting, & Demetriou, 2000; Piaget, 1932/1997; see Smetana, 2006). Thus, when in such interactions, children may display a more advanced command of these concepts and at still younger ages than we have shown here. Alternatively, our use of child–adult interactions in a laboratory may have made children comply with what they perceived as the adults' expectations. Note, though, that our helping measure did not entail providing responses about the moral transgressions directly to an adult; rather, it was a more implicit assessment that took place after, and in a different situation than, the transgressions. Children's responses are thus unlikely to have been influenced much by their perceptions of the adults' expectations. Still, on the view that children were complying with such expectations, children may not show the same sensitivity in interactions with other children. Even so, our findings at least demonstrate young children's ability to recognize harmful intentions in interactions with adults under controlled conditions. It remains for future research to assess this ability across multiple contexts and types of interactions.

Relatedly, although we assessed the impact of the intentions behind and outcomes of moral behavior, we recognize that these are only two of several criteria that fall under only one of several domains that impact children's moral judgments. Other criteria in the moral domain include, for instance, whether the perpetrator apologized and whether she was already punished (e.g., Miller & McCann, 1979). Criteria in the psychological domain include the victim's and the children's own relationship with the perpetrator (e.g., Slomkowski & Killen, 1992; Wellman et al., 1979). The context of the transgression and individual differences among children also play a role, and of course, all of the aforesaid domains and factors interact with each other (see Helwig, 2006; Killen, 2007; Smetana, 2006, for reviews). Intentions and outcomes thus form only one piece of the rich and multifaceted area of children's moral judgments.

Furthermore, we broadly used the term *moral behavior* throughout to mean acts that have consequences for others' rights or welfare (e.g., Smetana, Schlagman, & Adams, 1993). Morality certainly includes other criteria, such as obligation and independence from authority sanctions, which distinguish it from the conventional, psychological, and personal domains of social knowledge (e.g., Smetana, 2006; Turiel et al., 1987). Our focus here, however, was on whether children's prosocial behavior is mediated by an individual's harmful transgressions against others. Although these are precisely the types of transgressions that fall into the moral domain, our focus on prosocial behavior rather than on moral judgments precluded an assessment of whether children perceived the transgressions as moral in the strict sense or not. Children may have withheld help for many reasons (such as how much they liked the perpetrator) that are related to but not the same as judging the transgressions to be moral. All the same, children's ability to recognize harmful behaviors and intentions as seen in their prosocial behavior might be an early step on the way to the full-blown, explicit moral judgments that children make just a few years later and as such, is important to explore.

Finally, to say that children's prosocial behavior is mediated by others' harmful behaviors and intentions is not to disregard the many additional mediating factors. For instance, there is evidence for a direct reciprocity effect on the sharing behavior of 3-year-olds (Levitt, Weber, Clark, & McDonnell, 1985; Olson & Spelke, 2008). Sympathy for the potential beneficiary also plays a role (see Eisenberg & Miller, 1987; Vaish et al., 2009). Overall, the emerging picture is that early in development, children begin engaging in sophisticated social and moral evaluations that impact their own prosocial behavior.

References

Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attribution biases. *Journal of Experimental Child Psychology*, *55*, 177–193.

Arsenio, W., & Lover, A. (1995). Children's conceptions of sociomoral affect: Happy victimizers, mixed emotions and other expectancies. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 87–128). Cambridge, UK: Cambridge University Press.

Astington, J. W. (2004). Bridging the gap between theory of mind and moral reasoning. *New Directions for Child and Adolescent Development*, *103*, 63–72.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Cushman, F. A. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly*, *28*, 347–367.

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556.

Dunn, J., Cutting, A. O., & Demetriou, H. (2000). Moral sensibility, understanding others, and children's friendship interactions in the preschool period. *British Journal of Developmental Psychology*, *18*, 159–177.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*, 91–119.

Eisenberg-Berg, N., & Hand, M. (1979). The relationship of preschoolers' reasoning about prosocial moral conflicts to prosocial behavior. *Child Development, 50*, 356–363.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785–791.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(22), 557–560.

Helwig, C. C. (2006). Rights, civil liberties, and democracy across cultures. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 185–210). Mahwah, NJ: Erlbaum.

Hill, K., & Hill, C. (1977). Children's concepts of good and bad behavior. *Psychological Reports, 41*, 955–958.

Imamoğlu, E. O. (1975). Children's awareness and usage of intention cues. *Child Development, 46*, 39–45.

Janssens, J. M. A. M., & Dekoviç, M. (1997). Child rearing, prosocial moral reasoning, and prosocial behaviour. *International Journal of Behavioral Development, 20*, 509–527.

Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychological Bulletin, 85*, 76–85.

Killen, M. (1991). Social and moral development in early childhood. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 115–138). Hillsdale, NJ: Erlbaum.

Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science, 16*, 32–36.

Krebs, D. L. (2008). Morality: An evolutionary account. *Perspectives on Psychological Science, 3*, 149–172.

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect. *Psychological Science, 17*, 421–427.

Levitt, M. J., Weber, R. A., Clark, M. C., & McDonnell, P. (1985). Reciprocity of exchange in toddler sharing behavior. *Developmental Psychology, 21*, 122–123.

Miller, D. T., & McCann, C. D. (1979). Children's reactions to the perpetrators and victims of injustices. *Child Development, 50*, 861–868.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition, 108*, 222–231.

Piaget, J. (1997). *The moral judgment of the child*. New York: Free Press. (Original work published 1932)

Rhine, R. J., Hill, S. J., & Wandruff, S. E. (1967). Evaluative responses of preschool children. *Child Development, 38*, 1035–1042.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*, 466–469.

Slomkowski, C. L., & Killen, M. (1992). Young children's conceptions of transgressions with friends and nonfriends. *International Journal of Behavioral Development, 15*, 247–258.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.

Smetana, J. G., Schlagman, N., & Adams, P. W. (1993). Preschool children's judgments about hypothetical and actual transgressions. *Child Development, 64*, 202–214.

Turiel, E. (2006). Thought, emotions, and social interactional processes in moral development. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 7–35). Mahwah, NJ: Erlbaum.

Turiel, E., Killen, M., & Helwig, C. (1987). Morality: Its structure, functions, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155–243). Chicago: University of Chicago Press.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective-taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45*, 534–543.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383–403.

Vaish, A., Missana, M., & Tomasello, M. (in press). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*.

Wainryb, C., Brehl, B., & Matwin, S. (2005). Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal activities. *Monographs of the Society for Research in Child Development, 70* (Serial No. 281).

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*, 1301–1303.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*, 271–294.

Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology, 100*, 455–471.

Weiner, B., & Peter, N. (1973). A cognitive-developmental analysis of achievement and moral judgments. *Developmental Psychology, 9*, 290–309.

Wellman, H. M., Larkey, C., & Somerville, S. C. (1979). The early development of moral criteria. *Child Development, 50*, 869–873.

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development, 67*, 2478–2492.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1670–1677

Prospective Associations Between Peer Victimization and Aggression

Jamie M. Ostrov
University at Buffalo, The State University of New York

The current study involved a short-term longitudinal study of young children (M = 44.56 months, SD = 11.88, N = 103) to test the prospective associations between peer victimization and aggression subtypes. Path analyses documented that teacher-reported physical victimization was uniquely associated with increases in observed physical aggression over time. The path model also revealed that teacher-reported relational victimization was uniquely associated with statistically significant increases in observed relational aggression over time. Ways in which these findings extend the extant developmental literature are discussed.

The study of peer victimization has become an important focus for psychologists who attempt to understand the etiology, course, and outcomes associated with various forms of psychopathology including aggressive behavior in children (e.g., Hanish & Guerra, 2000; Hodges & Perry, 1999; Juvonen & Graham, 2001; Schwartz, McFadyen-Ketchum, Dodge, Pettit, & Bates, 1998; Troop-Gordon & Ladd, 2005). Rather than focusing on the developmental antecedents of peer victimization, the present study focuses on aggression as an important hypothesized outcome associated with peer victimization. Importantly, peer victimization (i.e., chronic or frequent receipt of aggression from peers over time) is associated with increases in physical aggression (i.e., the use or threat of physical force to hurt, harm, or injure another person; Dodge, Coie, & Lynam, 2006) over time in school-aged samples (Graham & Juvonen, 1998; Hanish & Guerra, 2000, 2002; Hodges, Boivin, Vitaro, & Bukowski, 1999; Khatri, Kupersmidt, & Patterson, 2000; Kochenderfer-Ladd, 2003; Schwartz et al., 1998). In fact, Hanish and Guerra (2002) proposed that peer victimization might be a risk factor for subsequent serious adjustment issues (e.g., externalizing behavior) because children who are victimized may have limited opportunities for peer interactions and developmentally salient socialization experiences. Thus, the focus of the present manuscript is testing if peer victimization predicts increases in aggressive behavior for young children. To further test this fundamental question we also examine if the developmental paths from victimization to aggression are present for both physical and relational victimization in early childhood.

Theories have been posited that relational victimization may be associated with increases in relationally aggressive behaviors in which the relationship serves as the means of harm (e.g., threatening friendship withdrawal or direct social exclusion; Crick & Grotpeter, 1996). Two additional conceptually related but distinct forms of aggression include indirect (Lagerspetz, Björkqvist, & Peltonen, 1988; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003) and social (Galen & Underwood, 1997) aggression (see Card, Stucky, Sawalani, & Little, 2008), which were not the focus of the current study. Rose and Rudolph (2006) present a peer-socialization model and are the first known theorists to articulate a conceptual model with potential paths from physical *and* relational peer victimization (i.e., ''exposure to peer stress'') to aggressive behavior (p. 116). According to this model, for example, ''greater exposure to overt or physical victimization may contribute to their vulnerability . . . to aggression. Physical aggression is more appropriate in response to overt victimization than subtle [relational] forms of victimization'' (Rose & Rudolph, 2006, p. 120). In addition, Yeung and Leadbeater (2007, p. 980) asserted that relationally victimized children may respond with relationally aggressive behaviors to hurt the aggressor or avoid further attacks.

The contributions of Emily E. Ries, Jamie L. Guzzo, Kirstin Stauffacher Grös, Stephanie A. Godleski, and the entire UB Social Development Lab staff are greatly appreciated. The author thanks Drs. Julie C. Bowker, Craig R. Colder, and Leonard J. Simms for statistical consultation and feedback on an earlier draft of this manuscript; he also thanks Emily J. Hart and Adam D. Mullins for their comments and suggestions and acknowledges the families, teachers, and directors of participating schools. Preparation of this manuscript was facilitated by a grant (R03HD059781) to the author from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Correspondence concerning this article should be addressed to Jamie M. Ostrov, 214 Park Hall, Department of Psychology, University at Buffalo, Buffalo, NY 14260-4110. Electronic mail may be sent to jostrov@buffalo.edu.

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0004

Collectively, the contentions that children retaliate in kind (i.e., physical victimization is uniquely associated with physical aggression and relational victimization is uniquely associated with relational aggression) have been articulated as the *specificity hypothesis of aggression* (Crick, Casas, & Ku, 1999; Ostrov, 2008). Young children are posited to be influenced primarily by their own personal experiences (i.e., history of adaptation) of peer victimization and aggression, but they may learn by being a bystander or uninvolved observer of peer interactions if they identify (i.e., label the behaviors as consistent with their self-construal) with the experiences of the salient model (i.e., perpetrator or victim). Thus, a child who is victimized may learn from these interactions and if reinforced will model these self-relevant behaviors increasingly in future hostile interactions with peers. Prior concurrent studies have found support for this hypothesis (Crick et al., 1999; Sullivan, Farrell, & Kliewer, 2006) or limited support depending on the method and informant (Putallaz et al., 2007), but no known prospective analysis has examined the links between both physical and relational victimization and changes in physical and relational aggression, respectively. Thus, the primary goal of the current study was to test if peer victimization uniquely predicted increases in aggression subtypes.

There are only a few known concurrent studies that have studied both aggression subtypes and victimization in early childhood, which is an important developmental period for understanding the onset and subsequent course of peer victimization (e.g., Bonica, Arnold, Fisher, Zeljo, & Yershova, 2003; Crick et al., 1999; Dhami, Hoglund, Leadbeater, & Boone, 2005). Prospective longitudinal studies exploring the association between peer victimization and future aggression subtypes are rare in the extant developmental literature (for exceptions in middle childhood and early adolescence, see Leadbeater & Hoglund, 2009; Schwartz et al., 1998). One recent study, in early adolescence, provided evidence that self-reported relational victimization was significantly associated with future self-reported relational aggression, even after controlling for physical victimization (Yeung & Leadbeater, 2007). To avoid shared-method bias concerns, Yeung and Leadbeater (2007) called for multi-informant sources of data in future work testing these links (see also Prinstein, Boergers, & Vernberg, 2001). As such, the current multi-informant and multimethod study (i.e., teacher report and naturalistic observations) advances the developmental literature by being the first to study how teacher-reported peer victimization subtypes may predict changes in observed aggressive behavior subtypes for young children.

Peer rejection may affect aggression and victimization status (Bierman, 2004; Bierman & Wargo, 1995; Perry, Kusel, & Perry, 1988), and thus, it is important to consider in any peer victimization study. Importantly, peer victimization and rejection are positively related across time, but they are independent constructs (Hodges & Perry, 1999; Perry et al., 1988). Both physical and relational peer victimization have been found to increase a child's likelihood of peer rejection (e.g., Crick & Bigbee, 1998; Hanish & Guerra, 2002), and peer rejection has been associated with increases in aggression subtypes over time (e.g., Werner & Crick, 2004). Aggression and peer victimization are related to peer rejection, but they are often found to be associated with peer rejection for different reasons (Boivin & Hymel, 1997; Perry et al., 1988). Given the moderate to high associations between peer victimization and rejection it is important to test models that include both constructs so that unique associations may be tested (see Morrow, Hubbard, Rubin, & McAuliffe, 2008).

In the present study, aggression is defined as behaviors in which there was the intent to hurt, harm, or injure another person (Dodge et al., 2006). It is well documented in the literature that there are other behaviors that resemble physical and relational aggression, but they do not conform to the definition of aggression. Developmental scholars have called these behaviors: "rough and tumble play," "assertion," or "social dominance," because they explicitly do *not* include the intent to harm others (e.g., Ostrov, Pilat, & Crick, 2006; Pellegrini, 2003). Moreover, exclusion behaviors that are considered fair (e.g., excluding a poor athlete from a sports group; Killen & Stangor, 2001) or actions in which the intent of the child is due only to a desire to play by him- or herself are not examples of relational aggression and were explicitly addressed in training of observers in the present study. It is important to note that children during early childhood (i.e., 3- to 5-year-olds) are able to engage in relatively sophisticated forms of relational aggression (see Crick, Ostrov, & Kawabata, 2007). However, during early childhood, even when these behaviors resemble sophisticated covert actions like spreading a malicious secret, they are direct in nature, based on situations occurring in the present, and the identity of the perpetrator is clearly defined (see Ostrov, Woods, Jansen, Casas, & Crick, 2004).

There were several goals and hypotheses for this study. Our central goal was to test the direct prospective associations between peer victimization subtypes and increases in aggressive behaviors. Based on the adopted conceptual model (Rose & Rudolph, 2006) and the specificity hypothesis of aggression (Crick et al., 1999; Ostrov, 2008), we predicted that peer victimization would be associated with increases in aggressive behavior. More specifically, in keeping with theory, we predicted that physical victimization would be positively and uniquely associated with increases in physical aggression (e.g., Hanish & Guerra, 2002). Further, we anticipated that relational victimization would be positively and uniquely associated with increases in relational aggression (Yeung & Leadbeater, 2007).

Method

Participants

A total of 119 (69 girls) children initially participated in this ongoing longitudinal study (e.g., Ostrov, 2008). Attrition was small (13%, 7 boys and 9 girls) and was because of families who moved out of the area, which yielded a final sample size of 103 children (60 girls). Children were on average 44.56 months old ($SD = 11.88$) at the first time point. Children represented various ethnic backgrounds: 14.6% African American, 14.6% Asian, 59.1% Caucasian, 2.9% Hispanic, 3.9% Indian, 1.0% Native American, and 3.9% other ethnicities. Most children were from middle-class families. Children were recruited in two equal cohorts and there were no statistically significant cohort effects for any of the key study variables ($t = -1.57$–1.19, $p > .12$).

Procedure

The study was approved by the local Institutional Review Board and parents provided written consent prior to participation. Head teachers also provided written informed consent (100% participating) prior to completing teacher reports. The second time point was approximately 4 months after the conclusion of Time 1 (for additional details, see Murray-Close & Ostrov, 2009).

Measures

Observations of aggression. The Early Childhood Observational System introduced by Ostrov and Keating (2004) and revised by Crick et al. (2006) was adopted (for a review, see Leff & Lakin, 2005). Naturalistic focal child sampling with continuous recording of children's relational and physical aggression were conducted during free play at participating child-care centers at both time points. At each of the two time points, every focal child was observed for eight separate sessions (10 min per session), which took place on different days over a 2-month period for 80 min of total observation per child.

Training consisted of several steps (see Ostrov, 2008). Assessments of reliability were conducted throughout the study to avoid observer drift (Pellegrini, 2004). For each 10-min session, the observers recorded, in full detail using a structured observation form: (a) physical aggression (e.g., hitting, kicking, pushing) and (b) relational aggression (e.g., direct verbal [i.e., ''You can't play with me''] or nonverbal exclusion from a group or dyadic interaction [i.e., blocking a child from entering a preferred area], using friendship withdrawal as a threat [i.e., ''You are not my friend anymore'']). Behaviors were summed within category, across the eight sessions, to yield total scores. Interobserver reliability (15% of observations) was acceptable for relational aggression (intraclass correlation coefficients [ICCs] from .72 to .86) and physical aggression (ICCs from .78 to .95). Correlations between teacher-reports and the observations of relational and physical aggression (at both time points) were significant and moderate in nature (see Ostrov, Ries, Stauffacher, Godleski, & Mullins, 2008).

Teacher report of victimization. The Preschool Peer Victimization Measure–Teacher Report (PPVM–TR; Crick et al., 1999) assessed physical and relational victimization at the first time point. This measure includes eight items: Two items assess physical victimization (''This child gets pushed or shoved by peers''), three items assess relational victimization (''This child gets left out of the group when someone is mad at them or wants to get back at them''), and three filler items assess received prosocial behavior. Teachers rated how often focal children were relationally or physically victimized on a 5-point scale ranging from 1 (*never to almost never true*) to 5 (*always or almost always true*). The items were summed within category. In the current study, Cronbach's αs = .85 and .90 for relational and physical victimization, respectively. Moderate levels of stability were found for both subscales, and in this sample the validity of the measure has been supported by replicating the original predicted factor structure and by obtaining significant correlations between teacher and research assistant (RA) reports for both subtypes of victimization (Ostrov, 2008).

Teacher report of peer rejection. The Preschool Social Behavior Scale–Teacher Form (PSBS–TF; Crick, Casas, & Mosher, 1997) was used to assess peer rejection (e.g., Crick et al., 2006). Using the aforementioned 5-point scale, teachers reported how disliked children were by same and opposite-sex peers (two summed items). This measure had acceptable reliability in the past (e.g., Crick et al., 2006) and in the current study was internally consistent at both times (Cronbach's $\alpha > .88$). Teacher-reported peer rejection was moderately stable across the study and teacher- and RA-reported peer rejection were significantly correlated (Ostrov, 2008).

Results

We first conducted descriptive statistics for the study variables (see Table 1). Two outliers (> 3 *SD*) were detected for observed relational aggression at Time 2 and these values were replaced with the value of 3 *SD* above the mean (Kline, 1998). Skew was less than 2 and kurtosis was lower than 8 for all variables, suggesting that non-normality of the data was not a concern (Kline, 1998). We also conducted correlations between the study variables (see Table 1). To avoid being underpowered for testing effects at the level of gender, only the total sample was used in this study. Importantly, physical victimization was significantly associated with future physical aggression ($r = .35$, $p < .001$) and relational victimization was significantly correlated with future relational aggression ($r = .27$, $p = .007$). Both physical and relational victimization were significantly correlated with peer rejection. Next, Fisher's r to Z tests were conducted to examine differences in the association between variables. The correlation between relational victimization at Time 1 and relational aggression at Time 2 ($r = .27$, $p = .007$) was significantly higher than the correlation between physical victimization at Time 1 and relational aggression at Time 2 ($r = .03$, *ns*), $Z = 2.21$, $p < .05$. It was also found that the association between peer rejection at Time 1 and relational victimization at Time 1 ($r = .56$, $p < .001$) was significantly higher than the corresponding association between peer rejection and initial physical victimization ($r = .24$, $p = .013$), $Z = 3.42$, $p < .01$.

Observational methods that rely on multiple observers and several sessions have often generated lower levels of intercorrelation between aggression subtypes than other approaches that may have higher levels of association as a result of shared method variance (see Ostrov & Crick, 2007). It is also conceivable that the amount of overlap even with teacher reports for 3- to 5-year-olds (e.g., .46 for boys and .37 for girls reported in Crick et al., 1997) is lower than for children in older developmental periods (e.g., .77; Crick, 1996) because physical aggression may be more common than relational aggression in this developmental period (see NICHD Early Child Care Research Network, 2004). In the present study, similar to past findings, the intercorrelation between observed physical and relational aggression was low and the intercorrelation between teacher-reported physical and relational victimization was moderate in magnitude (see Table 1).

To test the key hypothesis of the study a path analysis model with maximum likelihood estimation using Lisrel 8.54 was conducted (see Figure 1). In keeping with common practice in the field, in the tested models, age was entered as a covariate (e.g., Bonica et al., 2003; Crick et al., 2006; Hart, Nelson, Robinson, Olsen, & McNeilly-Choque, 1998). Models with and without gender as a covariate were virtually identical, and thus, the more parsimonious model without gender is presented. In keeping with recommendations of Hu and Bentler (1999) the standardized root mean square residual

Table 1
Descriptive Statistics and Correlations Between Key Study Variables

	1	2	3	4	5	6	7
Physical aggression OBS Time 1							
Relational aggression OBS Time 1	.18						
Physical aggression OBS Time 2	.46***	.15					
Relational aggression OBS Time 2	.08	.37***	.22*				
Peer rejection TR Time 1	.23*	.18	.28**	.05			
Physical victimization TR Time1	.34***	.001	.35***	.03	.24**		
Relational victimization TR Time1	.14	.29**	.18	.27**	.56***	.38***	
M	2.53	1.04	2.01	1.54	3.01	3.73	5.80
SD	2.54	1.55	2.42	2.12	1.47	1.84	2.55
Range	0–13	0–7	0–11	0–9	2–8	2–10	3–12

Note. OBS = observations; TR = Teacher report.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

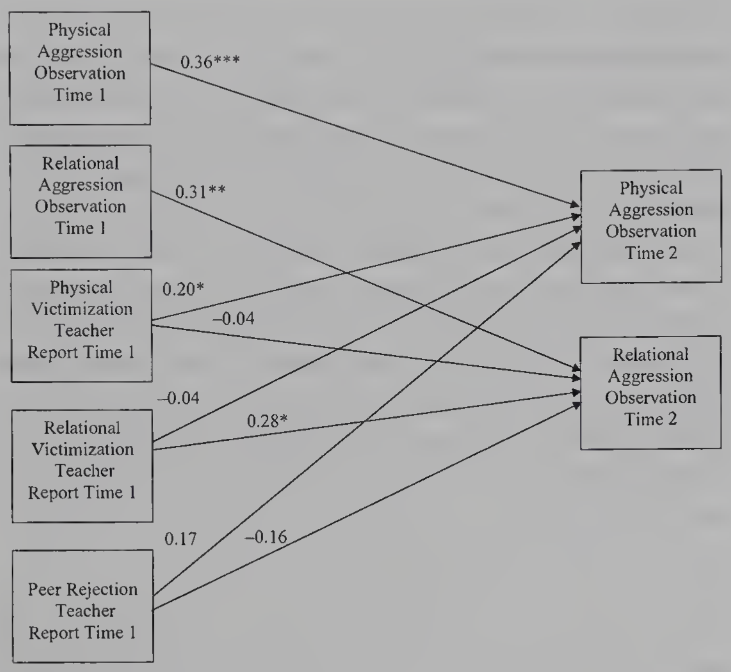

Physical Aggression Observation Time 1
Relational Aggression Observation Time 1
Physical Victimization Teacher Report Time 1
Relational Victimization Teacher Report Time 1
Peer Rejection Teacher Report Time 1
Physical Aggression Observation Time 2
Relational Aggression Observation Time 2
0.36***
0.31**
0.20*
−0.04
−0.04
0.28*
0.17
−0.16

Figure 1. Path analysis model predicting changes in observed physical and relational aggression.
Note. Standardized path coefficients are shown. Covariance paths (see text) and disturbances, as well as the covariate age, are not shown for ease of communication. The model fit was acceptable, $\chi^2(5) = 6.36$, $p = .27$; root mean square error of approximation = .053, standardized root mean square residual = .037, comparative fit index = .99.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

(SRMR) and the comparative fit index (CFI) were used to evaluate goodness of model fit. A cutoff of .08 or lower for SRMR and a value of .95 or larger for the CFI is viewed as a good fit (Hu & Bentler, 1999). The SRMR and CFI have been found to provide an appropriate balance between Type I and II error rates (Hu & Bentler, 1999). In addition, in keeping with standard reporting practices, the root mean square error of approximation (RMSEA) is also reported and acceptable values are considered less than .06 (Hu & Bentler, 1999).

The first model tested was the hypothesized model and the goodness-of-fit indices were acceptable, $\chi^2(5) = 6.36$, $p = .27$; RMSEA = .053, SRMR = .037, CFI = .99. The hypothesized model was compared with a more specified model but the additional two paths (i.e., between physical aggression at Time 1 and future relational aggression as well as from relational aggression at Time 1 to future physical aggression) were not significant. The model fit of this more specific model was generally adequate although the RMSEA was outside the acceptable range, $\chi^2(3) = 5.51$, $p = .14$; RMSEA = .093, SRMR = .030, CFI = .98. The $\Delta\chi^2$ test between the nested models was not significant, $\Delta\chi^2(2) = .85$, *ns*. In keeping with standard practice, the more parsimonious, hypothesized model was adopted. Importantly, the parameter estimates were similar and the same paths were significant in the two models. The adopted model also included (not shown in Figure 1 for ease of communication) a covariance between each of the predictor constructs and a covariance between the residuals of the outcome variables (i.e., physical and relational aggression at Time 2). We included initial levels of observed physical and relational aggression in the model so that testing increases in aggressive behavior over time could be analyzed. Importantly, as hypothesized, physical victimization was significantly associated with increases in physical aggression. In addition, as hypothesized, relational victimization was associated with significant increases in relational aggression (see Figure 1).

Discussion

The central study goals were to examine the prospective associations between peer victimization subtypes and future aggression subtypes. Rose and Rudolph (2006) posited that physical aggression rather than more subtle forms of behavior like relational aggression would be a more appropriate behavioral response to physical victimization. In keeping with the peer-socialization model (Rose & Rudolph, 2006) and past research (Hanish & Guerra, 2002), we found that peer physical victimization was associated with increases in physically aggressive behavior across time. Moreover, responding with the same behavior in kind has been posited for relational forms of behavior as well (see Yeung & Leadbeater, 2007). In keeping with these predictions and the *specificity hypothesis of aggression* (Crick et al., 1999; Ostrov, 2008), we found that teacher-reported relational victimization was associated with increases in observed relational aggression. Given the moderate and significant overlap between relational victimization and peer rejection, it is important to note that the conservative model also controlled for the influence of peer rejection (Morrow et al., 2008). Thus, it is not just that these children are rejected by their peers, which directly leads to changes in aggressive behavior; rather, the current findings suggest that experiencing victimization is a unique predictor of increases in aggression. Moreover, the present findings suggest that children who are victimized may learn from these interactions and adopt behaviors from salient social models (see Rose & Rudolph, 2006). Finally, if these actions are reinforced over time, presumably because they are effective at reducing peer victimization and facilitating goal obtainment such as social status, the child may continue to increase in his or her rate of aggression.

The present study addressed theoretically driven hypotheses with a multi-informant and multimethod short-term longitudinal study. Despite the strengths of this approach and use of time-intensive observational methods, the current study is limited in several ways. First, we were unable to explore moderation by gender analyses because we were underpowered for testing these effects and future work should examine models for both boys and girls with a larger and more diverse sample (see Card et al., 2008). Second, additional time points would have permitted an examination of growth trajectories in the development of both aggression and victimization and would have permitted more robust estimates of the present effects. Third, it is conceivable that earlier aggressive behavior may have predicted peer victimization, which in turn predicts increases in aggression. That is, physically (or relationally) aggressive children may receive more physical (or relational) retaliation which in turn, contributes to increases in his or her own aggression. Future longitudinal work with earlier assessments is needed to test these additional feasible pathways. Fourth, multiple methods (e.g., peer ratings or nominations, observations) for assessing victimization subtypes should be conducted in future studies (Kochenderfer & Ladd, 1997).

The current findings raise several implications for future research. The present study indicates that our past understanding of how victimization and aggression are linked across time may need to be revised. That is, a large amount of theory and research has focused exclusively on prospective links from aggression to victimization (for two exceptions, see Leadbeater & Hoglund, 2009; Schwartz et al., 1998). In fact, past social process models of peer harassment have posited that relatively stable behavioral tendencies like aggressive behavior may directly or indirectly lead to peer victimization via peer rejection (Boivin, Hymel, & Hodges, 2001). The current study used multiple methods and high levels of statistical control to examine the reverse direction of effect and found prospective links from peer physical victimization to physical aggression during early childhood, even with peer rejection in the model. We are also the first known study to reveal that relational victimization may uniquely lead to increases in relational aggression among young children. Thus, we may tentatively conclude a unique link between exposure to peer stress/harassment and increases in aggression over time (Rose & Rudolph, 2006). More specifically, we have demonstrated initial prospective support for the specificity hypothesis of aggression (Crick et al., 1999; Ostrov, 2008). Replication of the specificity hypothesis of aggression is needed with more diverse samples. In addition, the extent to which the children in the current study were provocative victims or aggressive victims (i.e., high in both aggression and victimization) with associated emotional and behavioral regulation problems is unknown and requires further investigation (see Schwartz, 2000; Schwartz, Dodge, Pettit, & Bates, 1997; Schwartz, Proctor, & Chien, 2001).

Several implications for future intervention and prevention efforts are raised by the current study. The findings suggest that children who are physically or relationally victimized by their peers may be at risk for displaying increases in school-based aggression over short periods of time. Interventions designed to reduce peer victimization in early childhood classrooms should be developed to minimize the likelihood of this future occurrence and should convey effective strategies for handling peer victimization other than retaliation (i.e., reframing social cognitions and social goals, notification of a teacher or caregiver). That is, school-based early childhood intervention efforts will likely be enhanced if programs explicitly address both peer victimization and aggression subtypes among peers (see Leadbeater, Hoglund, & Woods, 2003). Developmentally appropriate, early prevention would center on facilitating prosocial and friendship formation skills, while also reducing words and actions that cause harm (see Ostrov et al., 2009). In addition, these early childhood programs should be designed to foster a culture in which physical and relational aggression (e.g., malicious social exclusion) are not acceptable (see Harrist & Bradley, 2003). These programs should also provide children with effective tools to decode and distinguish harmful, intention acts from benign behaviors (e.g., exclusivity among close friends; see Leff et al., 2009).

In conclusion, the present study builds on prior theory and research to test novel empirical questions concerning the prospective associations between peer victimization and aggression. Specifically, the present study highlights the importance of understanding peer victimization to predict increases in aggressive behavior in young children.

References

Bierman, K. L. (2004). *Peer rejection: Developmental processes and intervention strategies*. New York: Guilford.

Bierman, K. L., & Wargo, J. (1995). Predicting the longitudinal course associated with aggressive-rejected, aggressive (non-rejected), and rejected (non-aggressive) status. *Development and Psychopathology, 7*, 669–682.

Boivin, M., & Hymel, S. (1997). Peer experiences and social self-perceptions: A sequential model. *Developmental Psychology, 33*, 135–145.

Boivin, M., Hymel, S., & Hodges, E. V. E. (2001). Toward a process view of peer pejection and harassment. In M. Putallaz & K. L. Bierman (Eds.), *Aggression, antisocial, behavior and violence among girls: A developmental perspective* (pp. 265–289). New York: Guilford.

Bonica, C., Arnold, D. H., Fisher, P. H., Zeljo, A., & Yershova, K. (2003). Relational aggression and language development in preschoolers. *Social Development, 12*, 551–562.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79*, 1185–1229.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327.

Crick, N. R., & Bigbee, M. A. (1998). Relational and overt forms of peer victimization: A multiinformant approach. *Journal of Consulting and Clinical Psychology, 66*, 337–347.

Crick, N. R., Casas, J. F., & Ku, H. C. (1999). Relational and physical forms of peer victimization in preschool. *Developmental Psychology, 35*, 376–385.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–587.

Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology, 8*, 367–380.

Crick, N. R., Ostrov, J. M., Burr, J. E., Jansen-Yeh, E. A., Cullerton-Sen, C., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology, 27*, 254–268.

Crick, N. R., Ostrov, J. M., & Kawabata, Y. (2007). Relational aggression and gender: An overview. In D. J. Flannery, I. Waldman, & A. Vazsonyi (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 245–259). Cambridge, UK: Cambridge University Press.

Dhami, M. K., Hoglund, W. L., Leadbeater, B. J., & Boone, E. (2005). Gender-lined risks for peer physical and relational victimization in the context of school-level poverty in first grade. *Social Development, 14*, 532–549.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 719–788). New York: Wiley.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*, 589–600.

Graham, S., & Juvonen, J. (1998). Self-blame and peer victimization in middle school: An attributional analysis. *Developmental Psychology, 34*, 587–599.

Hanish, L. D., & Guerra, N. G. (2000). Predictors of peer victimization among urban youth. *Social Development, 9*, 521–543.

Hanish, L. D., & Guerra, N. G. (2002). A longitudinal analysis of patterns of adjustment following peer victimization. *Development and Psychopathology, 14*, 69–89.

Harrist, A. W., & Bradley, K. D. (2003). "You can't say you can't play": Intervening in the process of social exclusion in the kindergarten classroom. *Early Childhood Research Quarterly, 18*, 185–205.

Hart, C. H., Nelson, D. A., Robinson, C. C., Olsen, S. F., & McNeilly-Choque, M. K. (1998). Overt and relational aggression in Russian nursery-school-age children: Parenting style and marital linkages. *Developmental Psychology, 34*, 687–697.

Hodges, E. V. E., Boivin, M., Vitaro, F., & Bukowski, W. M. (1999). The power of friendship: Protection against an escalating cycle of peer victimization. *Developmental Psychology, 35*, 94–101.

Hodges, E. V. E., & Perry, D. G. (1999). Personal and interpersonal antecedents and consequences of victimization by peers. *Journal of Personality and Social Psychology, 76*, 677–685.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for indexes in convariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Juvonen, J., & Graham, S. (2001). *Peer harassment in school: The plight of the vulnerable and victimized*. New York: Guilford.

Khatri, P., Kupersmidt, J. B., & Patterson, C. (2000). Aggression and peer victimization as predictors of self-reported behavioral and emotional adjustment. *Aggressive Behavior, 26*, 345–358.

Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development, 72*, 174–186.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Kochenderfer, B. J., & Ladd, G. W. (1997). Victimized children's responses to peers' aggression: Behaviors associated with reduced versus continued victimization. *Development and Psychopathology, 9*, 59–73.

Kochenderfer-Ladd, B. (2003). Identification of aggressive and asocial victims and the stability of their peer victimization. *Merrill-Palmer Quarterly, 49*, 401–425.

Lagerspetz, K. M. J., Björkqvist, K., & Peltonen, T. (1988). Is indirect aggression more typical of females? Gender differences in aggressiveness in 11- to 12-year-old children. *Aggressive Behavior, 14*, 403–414.

Leadbeater, B., & Hoglund, W. (2009). The effects of peer victimization and physical aggression on changes in internalizing from first to third grade. *Child Development, 80*, 843–859.

Leadbeater, B., Hoglund, W., & Woods, T. (2003). Changing contents? The effects of a primary prevention program on classroom levels of peer relational and physical victimization. *Journal of Community Psychology, 31*, 397–418.

Leff, S. S., Gullan, R. L., Paskewich, B. S., Abdul-Kabir, S., Jawad, A. B., Grossman, M., et al. (2009). An initial evaluation of a culturally-adapted social problem solving and relational aggression prevention program for urban African American relationally aggressive girls. *Journal of Prevention and Intervention in the Community, 37*, 260–274.

Leff, S. S., & Lakin, R. (2005). Playground-based observational systems: A review and implications for practitioners and researchers. *School Psychology Review, 34*, 475–489.

Morrow, M. T., Hubbard, J. A., Rubin, R. M., & McAuliffe, M. D. (2008). The relation between childhood aggression and depressive symptoms. *Merrill-Palmer Quarterly, 54*, 316–340.

Murray-Close, D., & Ostrov, J. M. (2009). A longitudinal study of forms and functions of aggressive behavior in early childhood. *Child Development, 80*, 828–842.

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood. *Monographs of the Society for Research in Child Development, 69* (Serial No. 278).

Ostrov, J. M. (2008). Forms of aggression and peer victimization during early childhood: A short-term longitudinal study. *Journal of Abnormal Child Psychology, 36*, 311–322.

Ostrov, J. M., & Crick, N. R. (2007). Forms and functions of aggression during early childhood: A short-term longitudinal study. *School Psychology Review, 36*, 22–43.

Ostrov, J. M., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development, 13*, 255–277.

Ostrov, J. M., Massetti, G. M., Stauffacher, K., Godleski, S. A., Hart, K. C., Karch, K. M., et al. (2009). An intervention for relational and physical aggression in early childhood: A preliminary study. *Early Childhood Research Quarterly, 24*, 15–28.

Ostrov, J. M., Pilat, M. M., & Crick, N. R. (2006). Assertion strategies and aggression during early childhood: A short-term longitudinal study. *Early Childhood Research Quarterly, 21*, 403–416.

Ostrov, J. M., Ries, E. E., Stauffacher, K., Godleski, S. A., & Mullins, A. D. (2008). Relational aggression, physical aggression and deception during early childhood: A multi-method, multi-informant short-term longitudinal study. *Journal of Clinical Child and Adolescent Psychology, 37*, 664–675.

Ostrov, J. M., Woods, K. E., Jansen, E. A., Casas, J. F., & Crick, N. R. (2004). An observational study of delivered and received aggression, gender, and social-psychological adjustment in preschool: "This white crayon doesn't work. . ." *Early Childhood Research Quarterly, 19*, 355–371.

Pellegrini, A. D. (2003). Perceptions and functions of play and real fighting in early adolescence. *Child Development, 74*, 1522–1533.

Pellegrini, A. D. (2004). *Observing children in their natural worlds: A methodological primer* (2nd ed.). Mahwah, NJ: Erlbaum.

Perry, D. G., Kusel, S. J., & Perry, L. C. (1988). Victims of peer aggression. *Developmental Psychology, 24*, 807–814.

Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. *Journal of Clinical Child Psychology, 30*, 479–491.

Putallaz, M., Grimes, C. L., Foster, K. J., Kupersmidt, J. B., Coie, J. D., & Dearing, K. (2007). Overt and relational aggression and victimization: Multiple perspectives within the school setting. *Journal of School Psychology, 45*, 523–547.

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132*, 98–131.

Schwartz, D. (2000). Subtypes of victims and aggressors in children's peer groups. *Journal of Abnormal Child Psychology, 28*, 181–192.

Schwartz, D., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1997). The early socialization of aggressive victims of bullying. *Child Development, 68*, 665–675.

Schwartz, D., McFadyen-Ketchum, S. A., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1998). Peer group victimization as a predictor of children's behavior problems at home and in school. *Development and Psychopathology, 10*, 87–99.

Schwartz, D., Proctor, L. J., & Chien, D. H. (2001). The aggressive victim of bullying: Emotional and behavioral dysregulation as a pathway to victimization by peers. In M. Putallaz & K. L. Bierman (Eds.), *Aggression, antisocial behavior, and violence among girls: A developmental perspective* (pp. 147–174). New York: Guilford.

Sullivan, T. N., Farrell, A. D., & Kliewer, W. (2006). Peer victimization in early adolescence: Associations between physical and relational victimization and drug use, aggression, and delinquent behaviors among urban middle school students. *Development and Psychopathology, 18*, 119–137.

Troop-Gordon, W., & Ladd, G. W. (2005). Trajectories of peer victimization and perceptions of the self and schoolmates: Precursors to internalizing and externalizing problems. *Child Development, 76*, 1072–1091.

Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? *Child Development, 74*, 1628–1638.

Werner, N. E., & Crick, N. R. (2004). Maladaptive peer relationships and the development of relational and physical aggression during middle childhood. *Social Development, 13*, 496–514.

Yeung, R. S., & Leadbeater, B. J. (2007). Does hostile attribution bias for relational provocations mediate the short-term association between relational victimization and aggression in preadolescence? *Journal of Youth and Adolescence, 36*, 973–983.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1678–1686

Development in Children's Comprehension of Linguistic Register

Laura Wagner
Ohio State University

Maia Greene-Havas
Towson University

Rebecca Gillespie
Xavier University

For socially appropriate communication, speakers must command a variety of linguistic styles, or *registers*, that vary according to social context and social relationships. This study examined preschool children's ability to use a speaker's register choice to infer the identity of their addressee. Four-year-olds could draw correct inferences based on Spanish and formal speech, and had limited success with infant-directed and casual styles. Five-year-olds drew appropriate inferences with all 4 styles; moreover, these children demonstrated strong explicit understanding of register, as measured through response justifications. These results demonstrate that children can use language for social judgments that extend beyond evaluating the speaker, and can interpret social implications of language even when not directly engaged in the particular social interaction.

We talk to different people differently. The speech style used to talk to a baby would be a strange way to address one's doctor and we do not talk to our employers the same way we talk to friends and peers. To engage in socially appropriate communication, we must command a variety of linguistic styles, or *registers*, that vary according to social context and our particular social relationship to the person we are speaking to. Part of acquiring a language, therefore, is learning what different linguistic registers signal about the social context as well as how to combine various linguistic elements into coherent register styles. The study reported here is an examination of children's developing skill at drawing inferences about the identity of a speaker's addressee based on the register that the speaker chooses to use.

Recent work has shown that young children are aware that language is a relevant indicator of social information. Kinzler, Dupoux, and Spelke (2007) found that 10-month-old infants prefer to take a toy from an adult that speaks their native language as opposed to a foreign language, and preschool-age children prefer to play with children who share their language natively, as opposed to speaking it with a foreign accent. Indeed, this social preference for people who speak the same way as oneself may even be stronger in some cases than a social preference for people of one's own race (Kinzler, Shutts, DeJesus, & Spelke, 2009). These results suggest that children understand how language is a marker of group affiliation; people who speak like you (or your parents) are part of your own group.

In many instances, language is in fact a stable trait of the speaker: people who are non-native speakers of English will generally remain that way. Certainly, they will not suddenly become native speakers of English just because they are given a new conversational partner. Native language (or native dialect) is therefore a good marker about stable properties of a speaker, such as where they are from. By contrast, linguistic register provides social information that shifts flexibly depending on the situation. The very same person speaks in one register when talking to an infant and a completely different register when speaking to her boss. Every individual must command a variety of linguistic registers to accommodate a variety of social situations. Identifying a specific register, therefore, tells one little about who the speaker is in general but

---

Special thanks are due to the Center for Science and Industry of Columbus, OH, where initial versions of this study were piloted. The data reported here served as the basis for the second author's honors thesis. Preliminary findings were presented at the International Association for the Study of Child Language conference held in Edinburgh, Scotland in 2008.

Correspondence concerning this article should be addressed to Laura Wagner, Department of Psychology, Ohio State University, 1835 Neil Ave., Columbus, OH 43210. Electronic mail may be sent to wagner.602@osu.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0005

instead indicates information about the social interaction the speaker is having.

Understanding register requires one to appreciate these social interactions, as well as the linguistic devices that are used to signal them. Registers are marked through a variety of linguistic elements. For example, infant-directed (ID) speech involves characteristic changes in pitch (Fernald, 1992), lexical items (e.g., ''goo goo''), and syntactic complexity (Newport, Gleitman, & Gleitman, 1977). Similarly, formal polite speech involves special lexical items (words like ''please'' and ''thank you''), syntactic constructions (Ervin-Tripp, 1977) and in some languages, pronoun choices (Brown, 1996). The specific lexical and syntactic elements used to signal different registers vary from language to language (Brown, 1996; Nakamura, 2001) and children must learn the mapping between linguistic elements and social situations. Moreover, there is also some variation in which social situations warrant distinctive registers (see e.g., Nakamura, 2001), so children must also learn their cultural norms to fully master this dimension of language use.

Previous work looking at children's ability to shift their own style of speech as a function of social context suggests that seeds of this ability may be present quite early: Weeks (1971) found that even a 1-year-old would shift his speech to reflect the social environment (e.g., indoor vs. outdoor speech). However, it is not until about age 4 years that children seem to command registers roughly parallel to those in the adult language. In a pioneering study, Shatz and Gelman (1973) found that 4-year-olds talk differently (e.g., using shorter, less complex utterances) to younger children than they do to either peers or adults. Moreover, in her extensive production studies, Andersen found that starting around age 4 years, children can shift their speaking style in response to a variety of social factors (Andersen, 1992, 1996; Andersen, Brizuela, DePuy, & Gonnerman, 1999). In these studies, children were given various puppets and asked to provide their voices in different situations. Children demonstrated sophisticated knowledge about the range of ways that language reflects social status. For example, when speaking for the doctor puppet, children used more imperative commands and asked fewer questions than when speaking for the nurse puppet. Similarly, children adjusted the choice of lexical items (using terms like ''sweetie'' more when speaking as the mommy puppet), the pitch of the utterance (using a deeper voice when speaking as the daddy puppet), and the use of discourse markers designed to help a speaker keep their turn in the conversation (they held their turn more aggressively when speaking as the doctor and teacher puppets).

Other researchers have similarly found that by age 4 years, children can shift their speech register to be more polite (James, 1978), and more impressively, to facilitate actual interpersonal interactions (Ervin-Tripp, 1977; Kyratzis, Marx, & Wade, 2001). However, children's skill at shifting registers improves dramatically with age. In Andersen's studies (Andersen, 1992, 1996; Andersen et al., 1999), the 6-year-olds were far more consistent about changing their speaking style and did so over a broader range of social contexts. Other researchers have also found that children improve their control of register throughout childhood. For example, children develop in their ability to ask ''nicely'' when requesting a desirable object from an older adult (Bates & Silvern, 1977; Nippold, Leonard, & Anastopoulos, 1982) and do not command the subtleties of polite speech until late in the grade school years (Pedlow, Sanson, & Wales, 2004; Pedlow, Wales, & Sanson, 2001). Children's comprehension of polite speech shows a similar developmental trajectory (Baroni & Axia, 1989; Bates & Silvern, 1977; Bernicot & Legros, 1987; Nippold et al., 1982; Pedlow et al., 2001; Pedlow et al., 2004).

This study expands on previous research in two critical ways. First and foremost, this study focuses on the intrinsic flexibility at the core of register: The very same individual should command multiple registers and deploy them appropriately depending on their listener. Previous experimental studies have examined only a single register at a time and therefore cannot speak to children's appreciation of how a single speaker shifts speech styles. This study tested children with three different registers, each associated with social interactions that should be reasonably familiar to children: formal polite speech (as a child might use to a teacher), casual speech (as a child would use to a peer), and ID speech (as one would speak to a baby). Second, this study focuses on children's comprehension as opposed to their production abilities. As noted previously, appropriate production of registers requires facility with a full range of linguistic levels and children may well understand the import of registers before they can produce an adult-like example of one. Indeed, recent work suggests that before 2 years of age, children can extract a great deal of information from speech not directly addressed to them, including detecting differences in fluency (Soderstrom & Morgan, 2007) and even learning the meanings of words (Akhtar, Jipson, &

Callanan, 2001; Floor & Akhtar, 2006). Similarly, children may be able to extract register information and use it despite limited command in their own production.

In addition, to provide a baseline of ability in this task, we tested children with a foreign language (Spanish). Hirschfeld and Gelman (1997) found that by age 4 years children associate people who speak a foreign language with foreign objects (e.g., non-Western housing and clothing). Linking a foreign language to someone from far away is conceptually similar to linking a particular register with the right social partner; in both cases, one needs to know that the language is providing information about the social identity of the listener. However, as even newborns can discriminate among different languages (Mehler et al., 1988), it was expected that the linguistic features of Spanish would be easy for children to detect and use.

In our task, children were asked to identify who a speaker was addressing based on her choice of register. In addition, for half the trials, we asked children to justify their responses. This additional task was added in an exploratory spirit to see if children's success with register was linked to their metalinguistic understanding of it. The motivation for this task came from the field of bilingualism. For bilingual speakers, the language they speak (e.g., French vs. Spanish) is not stable but instead, like register, switches flexibly depending on the social situation. Bilingual children tend to be quite sophisticated at understanding how social situations should guide their language choices (Petitto, Katerelos, Levy, & Tétreault, 2001) and they are ahead of their monolingual peers on some metalinguistic tasks (Bialystok, 2001; Jessner, 2006). Moreover, Jordá (2003) found that bilingual adults were better than monolinguals at providing relevant justifications for their performance on a linguistic task. To the extent that both register and bilingualism require one to choose among one's repertoire of speech styles to match a social situation, we might find a similar link between metalinguistic skill and register understanding.

Method

Participants

Participants were 56 children, divided into three groups: 3-year-olds ($n = 18$, $M = 42.2$ months), 4-year-olds ($n = 20$, $M = 55.5$ months), and 5-year-olds ($n = 18$, $M = 68.3$ months). Approximately half the children in each group were girls. The data from an additional 10 children were excluded because of equipment failure or experimenter error (5 children) or because the children failed to pass the pointing task described in the Procedures section (5 children, mean age = 46 months). All children were monolingual speakers of English from a large American Midwestern city; most of the children were Caucasian and were from a mid to high socioeconomic status. All were recruited by telephone and were tested in a university laboratory setting.

Stimuli

A frog puppet was the nominal speaker of all sentences. Neither the sex nor the age of the puppet was specified, and it spoke with the voice of a college-aged woman (the second author). All sentences were prerecorded and played as QuickTime files on a laptop computer.

The same four linguistic tokens were used throughout; each began with an attention-getting introduction and then asked for a person's name. Inquiring after names was chosen because pilot testing found that children accepted it as a plausible question to ask a variety of people, including infants. The tokens varied with respect to speech style and differed at many linguistic levels, including prosody, phonology, lexical choices, and syntax. Transcripts of the tokens can be found in Table 1. The ID token was spoken in a high-pitched voice with large pitch excursions; this special prosodic contour was particularly pronounced in the attention-getting word (''Awwww''). Syntactically, the ID token used an indirect question. The *formal* speech token used explicit politeness terms (''Excuse me,'' ''Please''), an indirect question in a modal verb construction (''can'') and did not contract across words phonologically. The *casual* token used an informal attention-getter (''Hey''), a simple question construction, and contracted the auxiliary verb. The *Spanish* token was spoken in a casual register in Spanish.

The target addressee for the ID speech was a picture of an infant; for the casual speech, a blonde toddler girl; for the formal speech, an adult woman; and for Spanish, a dark-haired toddler girl. Note that all the pictures (except those of the infants) depicted females. Thus, gender could not be used in the study to differentiate among the choices. The pictures were all presented on 8 in. × 6 in. laminated cards. There were three exemplars of each picture type used in the study. One example of each type is shown in Table 1.

Table 1
Linguistic and Visual Stimuli Used

Test register	Linguistic stimuli	Target picture	Target picture choice (1 example of 3 used)
Infant directed speech	*Awwww, I wonder what your name is?*	Baby	
Casual speech	*Hey, what's your name?*	Little girl who lives nearby	
Formal speech	*Excuse me please, can you tell me your name?*	Teacher	
Spanish	*Hola! Como se llamas?*	Little girl who lives far away	

Procedure

To insure that children were familiar with all the target pictures and could classify them appropriately, participants began by identifying the person in all 12 of the target pictures (3 of each target type) and sorting them into the four target groups. To ensure that children distinguished between the two kinds of little girls, the first example of each of these types was presented with a special introduction. For the little blonde girl, children were told: "This little girl lives right here in town. She lives very close by." For the little dark-haired girl, children were told: "This little girl lives very far away. She lives in a completely different country." Feedback was given during the sorting, and children were encouraged to use our labels for the pictures (baby, teacher, little girl who lives far away, little girl who lives nearby). These descriptions were repeated as needed throughout the sorting session. Following the sort, children were asked to point in a pseudo-random order to six pictures to insure that they could identify all the types. Children who failed to get 100% correct at this pointing task were excluded from analysis.

Children were then introduced to the frog puppet who wanted to meet new people and were asked to tell the frog their own name. They were told that the frog would meet some of the people in the pictures and were instructed to listen to the frog and say who she was talking to. For each test trial, children saw a pair of pictures: a target (the correct answer) and a foil. The foil pictures were all targets for different registers. The linguistic stimulus was played and the child was asked: "Who was she talking to?" Each type of target was tested three times, once paired with each of the possible foil picture types, for a total of 12 test trials. Children's responses consisted of selecting one of the two pictures. On the last six trials, children were also asked to justify their choices.

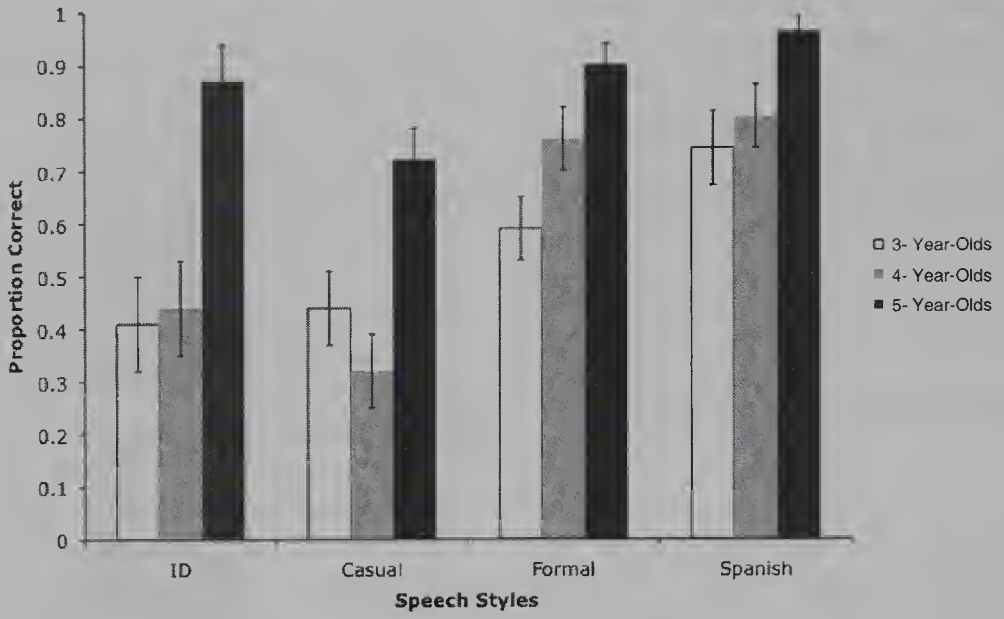


Figure 1. Mean correct responses by speech style and age.
Note. ID = infant directed.

Results

An analysis of variance (ANOVA) was conducted on the proportion correct with age as a between-subjects factor (3-, 4-, and 5-year-olds) and speech style as a within-subjects factor (ID, casual, formal, Spanish). The analysis found a main effect for age, $F(2, 53) = 24.3$, $p < .001$, $\eta^2 = .48$. Post hoc comparisons (Tukey's honestly significant difference [HSD], $p < .05$) found that 5-year-olds did significantly better than 4-year-olds who in turn did significantly better than 3-year-olds. A main effect was also found for speech style, $F(3, 51) = 9.5$, $p < .001$, $\eta^2 = .15$. Post hoc paired t tests (all ps < .05) found that children did significantly better with Spanish than all other speech styles, they did significantly better with the formal register than with the other two styles, and they did equally well with the casual register and the ID register. There was no significant interaction between age and speech style. The means for all registers by age group are shown in Figure 1.

To assess children's objective level of ability with each speech style, we compared their performance to chance (.50). The 5-year-olds were significantly above chance with all four speech styles: Spanish, $t(17) = 18.2$, $p < .001$, $d = 8.8$; formal, $t(17) = 11.2$, $p < .001$, $d = 5.4$; casual, $t(17) = 3.6$), $p < .01$, $d = 1.7$; and ID, $t(17) = 5.5$, $p < .001$, $d = 2.7$. Moreover, the number of 5-year-olds scoring at least .67 correct for all four speech styles was significantly above chance by a binomial test (17/18, $p < .001$); even for the hardest style (casual speech), a significant number of 5-year-olds were above chance (14/18, $p = .031$). The 4-year-olds were significantly above chance with Spanish, $t(19) = 5.3$, $p < .001$, $d = 2.4$ and the formal register, $t(19) = 4.2$, $p < .001$, $d = 1.9$, but performed at chance with the casual register, $t(19) = 0.74$, *ns*, and the ID register, $t(19) = 1.88$, *ns*. The subjects analysis confirmed that a significant number of participants performed above chance with Spanish and the formal register (in both cases, 17/20, $p = .003$ by binomial test), but not for ID speech (12/20, $p > .5$ by binomial test). However, despite the fact that the overall score was not above chance for the casual register, a significant number of 4-year-olds did score at least .67 correct with it (15/20, $p = .041$ by binomial test). Finally, the 3-year-olds scored no better than chance with the formal, casual, and ID registers (all ps > .1) but did score above chance with Spanish, $t(17) = 3.5$, $p = .003$, $d = 1.7$. However, the subjects analysis revealed that for none of the speech styles did a significant number of participants score above .67, not even for Spanish (13/18, $p = .096$ by binomial test). We were surprised at how poorly the children performed with ID speech and checked to see if children with more exposure to the use of ID speech—for example, children with younger siblings—would better understand its use. We compared the performance with ID speech of participants with and without younger siblings, but there was no difference between these groups, $F(1, 53) = 0.53$, $p > .4$. Raw exposure to ID speech directed to others did not determine success in understanding it.

Inspection of the data revealed that there were base-rate effects for choosing different types of

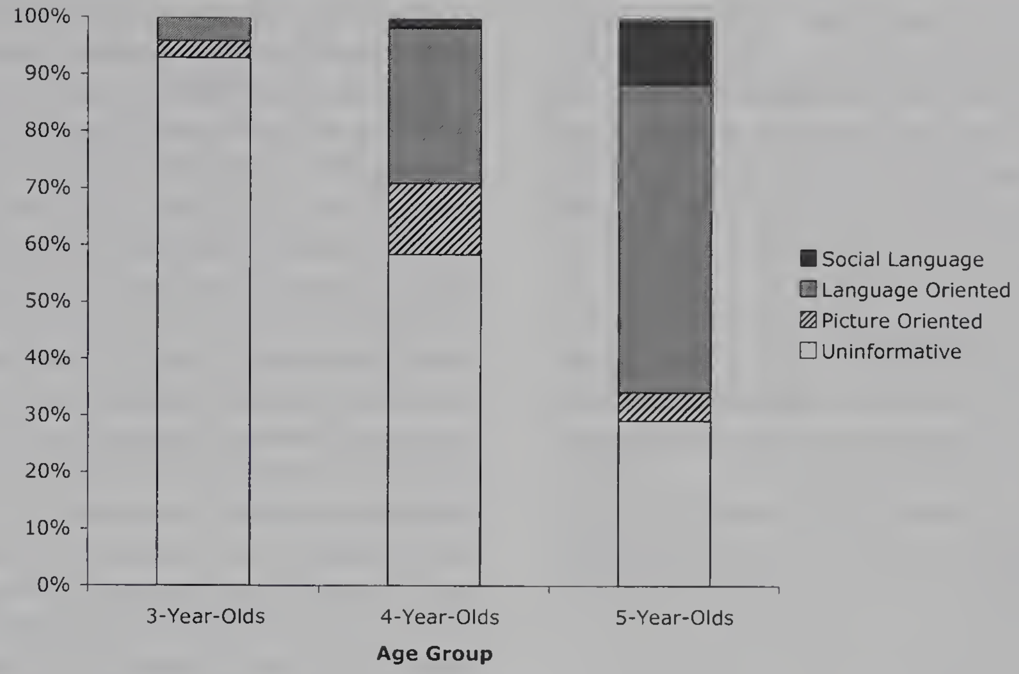


Figure 2. Percentage of the different justifications used.

pictures. In particular, 3- and 4-year-olds avoided choosing pictures of babies and were positively disposed to choose pictures of the girls who lived far away. The chance comparisons just noted may therefore have underestimated children's ability with ID speech (where the baby pictures were the correct targets) and overestimated their ability with Spanish (where the pictures of the girl who lived far away were the correct targets). Therefore, the chance analyses were redone with chance corrected to the actual base-rate for choosing the target (i.e., the average rate at which the target was chosen across the entire experiment). For the 5-year-olds, the base rates varied between .44 and .56 across targets, and the new analyses were qualitatively identical to the standard chance analysis. For the 4-year-olds, the base rates varied between .33 (for the baby pictures) and .63 (for the girls who lived far away). The corrected chance analyses yielded qualitatively identical results except for the case of ID speech: Four-year-olds did choose the picture of the baby significantly more frequently than their base rate preference, $t(19) = 2.2$, $p = .039$, $d = 1.0$. For the 3-year-olds, the base rates varied between .38 (for the baby pictures) and .66 (for the girls who lived far away). The corrected chance analyses also yielded qualitatively identical results except for the case of Spanish: Three-year-olds failed to choose the correct target more than their base rate preference, $t(17) = 1.2$, $p > .2$. One additional analysis (suggested by a reviewer) looked at whether the 3-year-olds could successfully contrast Spanish and the casual register just in those cases where the pictures contrasted the two kinds of little girls. The results found that they were equally likely to choose the little girl who lived far away with both those speech styles, $t(17) = 0.6$, *ns*, reinforcing the base-rate-corrected analysis that these children do not succeed with Spanish.

Finally, we analyzed children's justifications of their choices. Children's responses were classified into one of four categories. The first category was uninformative and was composed of justifications such as "I don't know" and "just because." The language-oriented category was broadly construed and included any answer making any reference to the linguistic stimuli, no matter how vague ("That's what it sounded like," "Because she had a pretty voice," "Because she said awwww"). It also included justifications that simply identified the stimuli as being Spanish ("Because she's speaking Spanish"). The picture-oriented category was used for references to the people in the pictures with no mention of language: "Because she had the same color hair as me," "Because she's from far away," Last, the social-language category consisted of answers that expressed an understanding of how the language was being tailored to the social category of the people in the pictures; for example, one child justified choosing the little girl who lives nearby after hearing the casual register by saying "because people say 'hi' to kids."

The distribution of the justifications is shown in Figure 2. The 3-year-olds' justifications were almost always uninformative (94.4%), which stands in marked contrast to the older children's justifications. One-way ANOVA comparisons between the 4- and 5-year-olds' justifications showed that the older children gave significantly fewer uninformative justifications, $F(1, 37) = 6.3$, $p = .017$, $\eta^2 = .15$, and significantly more language-oriented justifications, $F(1, 37) = 6.7$, $p = .014$, $\eta^2 = .16$, as well as social-language justifications, $F(1, 37) = 4.7$, $p = .036$, $\eta^2 = .12$. Moreover, children who gave better justifications also did better on the task as a whole. Looking just at the 4- and 5-year-old children (i.e., children who gave a range of justification types), a Pearson's correlation found a significant negative relation between children's overall proportion correct on the task and the proportion of uninformative justifications that they gave ($r = -.48$, $n = 38$, $p = .002$). That is, as children did better on this task, they were more likely to offer a reasonable justification for their choices.

Discussion

This study investigated children's developing ability to understand the social implications of using a particular register: Can a child tell who someone is speaking to based on the style in which they choose to talk? The results show that there is a dramatic improvement in children's ability to interpret register information around the age of 4 years and that this ability is broad based, applying flexibly to many different registers from a single speaker. Understanding of register continues to improve with age, and 5-year-olds are much more facile in their understanding. Moreover, these older children also have an improved appreciation of the basis of that understanding and provide substantially better justifications for their responses. These results suggest that there may be a connection between being flexible with register understanding and being able to explicitly think about language itself.

These results go beyond recent findings of Kinzler and colleagues (Kinzler et al., 2007; Kinzler et al., 2009) about children's ability to link linguistic and social information. Kinzler and colleagues found that preschool-age children can use the language of a speaker as a stable indicator of their social desirability; by contrast, these results show that by age 4 years, children also understand that speakers can shift their style of speaking depending on their conversational partner. Children's success with register suggests that they understand not only how language can stand proxy for stable social categories but also how language can be used to respond to different social partners. Moreover, the fact that children succeed with multiple registers shows that they understand the inherent flexibility of register; a speaker's social partners are subject to change and so are a speaker's registers.

The age of children's success was not notably different from what has been found in production studies (Andersen, 1992; James, 1978; Kyratzis et al., 2001; Shatz & Gelman, 1973), suggesting that children's comprehension of register is tied to their earliest successful attempts to produce it. What is notable about these results is the fact that the children succeeded despite the fact that they were essentially eavesdropping on other people's conversations. Akhtar and colleagues (Akhtar et al., 2001; Floor & Akhtar, 2006) have previously shown that children can listen in on adult conversations to learn the meanings of words, but the children in this study were extracting social linguistic information far beyond word meanings. Despite having no social position in the conversation of their own, the children were nevertheless identifying critical social characteristics of the participants (e.g., being a peer vs. being an adult) as well as social characteristics of the language being used (e.g., ID speech vs. formal speech). Children's success even as third-party observers shows how salient language is as a social marker (it is one that does not even require personal engagement); moreover, it points to a potential path for acquisition. Rogoff, Paradise, Arauz, Correa-Chávez, and Angelillo (2003) discuss the importance of eavesdropping—or what they call ''listening in''—in children's acquisition of a variety of skills cross-culturally. Interestingly, they argue that a critical feature to this process is that children anticipate that they will themselves eventually be asked to engage in the activity which they are listening in on. If this argument is correct, it implies that children may in fact anticipate the need to modulate their own speech style for different social settings before they themselves have direct need of it.

Finally, this study suggests that there is a positive link between children's ability in the comprehension task and their metalinguistic appreciation of the speech styles involved, as measured by their justifications. An examination of the differential success among the speech styles further supports the idea that the two are connected. For example, Spanish was the easiest style for children to link to a social category, and it was also the style for which children did the best job of justifying their choices: A full 47.9% of the language-oriented justifications made reference to the Spanish language in some way. No other speech styles were mentioned so explicitly in the justifications; however, a comparison between the formal and casual registers further suggests a role for explicit knowledge.

Children performed far better with the formal register than the casual one: Four-year-olds succeeded robustly with the former, but only marginally with the latter (although a significant number of 4-year-olds did comprehend this register successfully, children did not succeed as a group until age 5 years). The formal register was distinguished by its inclusion of polite terms such as "please" and "excuse me." Previous research suggests that in naturalistic settings, parents explicitly encourage their children to use these particular words when speaking to adults (Becker, 1994; Gleason, Perlmann, & Greif, 1984), and they appear to serve as critical markers of polite speech for children as old as 10 years (Pedlow et al., 2001). By contrast, the casual register was largely defined by what it lacked: It was normal, neutral speech with no distinctive intonational patterns or lexical items. The differential success of these two registers suggests that children may require specific, familiar linguistic markers to categorize the speech and connect it to their social understanding. If this suggestion is correct, it predicts that children may find it easier to understand register distinctions in languages which have specific grammatical elements associated with them (such as the way pronoun choice signals casual vs. formal speech in many Romance languages).

In both these cases, of course, the data is only correlational, and it is unclear whether children's facility with the speech styles is driving improvements in their metalinguistic abilities (presumably the direction of causation for bilingual children) or whether independent improvements in metalinguistic ability are helping them better comprehend speech registers. Further studies will be needed to determine if there is a strong parallel between knowing multiple languages and knowing multiple registers, or if the two are linked to improved metalinguistic understanding for different reasons.

Altogether, these results suggest that around age 4 years, children develop an appreciation for register and linguistic style that extends beyond their own social interactions. They understand that individual speakers can flexibly shift their speaking style as a function of their conversational partner; moreover, they understand which specific registers in their native language should be used with different social partners. Children's success reflects, on the one hand, a language specific ability, namely, the particular way that English signals formal, casual, and ID registers. The fact that there are some differences in how well children perform with these different registers speaks to the very English-specific knowledge that children must acquire. On the other hand, the fact that children succeed with multiple registers at once suggests that they have grasped the general principle underlying register use, namely, that language can be used to signal a social relationship between conversational partners.

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*, 416–430.

Andersen, E. S. (1992). *Speaking with style*. New York: Routledge.

Andersen, E.S. (1996). A cross-cultural study of children's register knowledge. In D. I. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 125–142). Mahwah, NJ: Erlbaum.

Andersen, E. S., Brizuela, M., DePuy, B., & Gonnerman, L. (1999). Cross-linguistic evidence of the early acquisition of discourse markers as register variables. *Journal of Pragmatics*, *31*, 1339–1351.

Baroni, M. R., & Axia, G. (1989). Children's meta-pragmatic abilities and the identification of polite and impolite requests. *First Language*, *9*, 285–297.

Bates, E., & Silvern, L. (1977). Social adjustment and politeness in preschoolers. *Journal of Communication*, *27*, 104–111.

Becker, J. (1994). Pragmatic socialization: Parental input to preschoolers. *Discourse Processes*, *17*, 131–148.

Bernicot, J., & Legros, S. (1987). Direct and indirect directives: What do young children understand? *Journal of Experimental Child Psychology*, *43*, 346–358.

Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics*, *21*, 169–181.

Brown, R. (1996). The language of social relationship. In D. I. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 39–52). Mahwah, NJ: Erlbaum.

Ervin-Tripp, S. (1977). Wait for me, Roller Skate! In S. Ervin-Tripp & C. Mitchell-Kernan (Eds.), *Child discourse* (pp. 165–188). New York: Academic Press.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow et al. (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, UK: Oxford University Press.

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*, 327–329.

Gleason, J. B., Perlmann, R. Y., & Greif, E. B. (1984). What's the magic word: Learning language through politeness routines. *Discourse Processes*, *7*, 493–502.

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development, 12*, 213–238.

James, S. L. (1978). The effect of listener age and situation on the politeness of children's directives. *Journal of Psycholinguistic Research, 7*, 307–317.

Jessner, U. (2006). *Linguistic awareness in multilinguals*. Edinburgh, Scotland: University of Edinburgh Press.

Jordá, M. P. S. (2003). Metapragmatic awareness and pragmatic production of third language learners of English: A focus on request acts realizations. *International Journal of Bilingualism, 7*, 43–69.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 12577–12580.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*, 623–634.

Kyratzis, A., Marx, T., & Wade, R. E. (2001). Preschoolers communicative competence: register shift in the marking of power in different contexts of friendship group talk. *First Language, 21*, 387–429.

Mehler, J., Juscyzk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition I young infants. *Cognition, 29*, 144–178.

Nakamura, K. (2001). The acquisition of polite language by Japanese children. In K. E. Nelson, A. Aksu-Koç, & C. E. Johnson (Eds.), *Children's language* (Vol. 10, pp. 93–112). Mahwah, NJ: Erlbaum.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 109–150). Cambridge, UK: Cambridge University Press.

Nippold, M. A., Leonard, L. B., & Anastopoulos, A. (1982). Development in the use and understanding of polite forms in children. *Journal of Speech and Hearing Research, 25*, 193–202.

Pedlow, R., Sanson, A., & Wales, R. (2004). Children's production and comprehension of politeness in requests: Relationships to behavioural adjustment, temperament, and empathy. *First Language, 24*, 347–367.

Pedlow, R., Wales, R., & Sanson, A. (2001). Children's production and comprehension of politeness in requests: Relationships to behavior adjustment in middle childhood. *Journal of Language and Social Psychology, 20*, 23–60.

Petitto, L. A., Katerelos, M., Levy, B. G., & Tétreault, K. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language, 28*, 453–496.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology, 54*, 175–203.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children ad a function of listener. *Monographs of the Society for Research in Child Development, 38*(5), 1–38.

Soderstrom, M., & Morgan, J. L. (2007). Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science, 10*, 641–653.

Weeks, T. E. (1971). Speech registers in young children. *Child Development, 42*, 1119–1131.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1687–1695

Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children

Stuart Marcovitch and
Janet J. Boseovski
University of North Carolina at Greensboro

Robin J. Knapp
University of Miami

Michael J. Kane
University of North Carolina at Greensboro

Goal neglect is the phenomenon of failing to execute the momentary demands of a task despite understanding and being able to recall the task instructions. Successful goal maintenance is more likely to occur in adults with high working memory capacity (WMC) who can keep rules mentally accessible while performing the task. The current study predicted that goal neglect would also be related to WMC in children. It assessed thirty-seven 4-year-old and twenty-eight 6-year-old children on the goal neglect version of the Dimensional Change Card Sort, and 3 tasks that measure WMC. As predicted, children with higher WMC scores were more likely to maintain goals adequately for task performance. The findings are consistent with a 2-factor model of working memory and its development.

Goal-directed behavior is complex, as it necessitates a representation of the goal state, formulation of a plan to achieve the goal state, and the ability to hold the goal in mind so that it may guide behavior at the appropriate time. Failure to behave accordingly, despite demonstrating knowledge of the appropriate actions, has been termed ''goal neglect'' (Duncan, Emslie, Williams, Johnson, & Freer, 1996) and has often been attributed to momentary lapses in working memory in typically developing populations (Kane & Engle, 2003) and associated with fluid intelligence and frontal lobe functioning (Duncan, 1995). Importantly, goal maintenance relies on self-regulatory skills such as persistence and resistance to distractibility, which are known to predict school readiness and academic success (Blair, 2002; McClelland et al., 2007).

We take the view that goal neglect can result in perseverative responding, the tendency to behave incorrectly in a manner that was previously correct. Perseverative responding is seen widely in infancy and early childhood and may result from the inability to maintain the goal representation sufficiently to override a prepotent response (Marcovitch & Zelazo, 2009). This view is consistent with a number of theories of goal-directed behavior in children, and we will illustrate its potency by examining theories of performance on the Dimensional Change Card Sorting task (DCCS; Zelazo, Frye, & Rapus, 1996; Zelazo, Müller, Frye, & Marcovitch, 2003), a popular measure of executive function in young children. In the DCCS, children are told an explicit rule to match cards according to one dimension (e.g., color). After several trials, the rule is changed so that the children now match cards by the other dimension (e.g., shape). For example, children may be instructed to match blue circles and green squares to a blue square and a green circle. They initially are instructed to sort by color, thus matching blue circles to the blue square and green squares to the green circle. After a number of trials, the rule is switched and they are instructed to sort by shape, thus matching blue circles to the green circle and green squares to the blue square. Although 3-year-olds can sort the cards by each rule independently, they have difficulty following the second rule after sorting by the first. Of importance, despite sorting incorrectly after hearing the second rule, children display knowledge of the rule when asked to point to the correct location (Zelazo

---

The authors thank G. Gleason for assisting with data collection and T. Kwapil for helpful discussions on statistical analyses.

Correspondence concerning this article should be addressed to Stuart Marcovitch, Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27402. Electronic mail may be sent to s_marcov@uncg.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0006

et al., 1996), when judging the sorting behavior of a puppet (Jacques, Zelazo, Kirkham, & Semcesen, 1999), and when told to verbalize it (Kirkham, Cruess, & Diamond, 2003; Towse, Redbond, Houston-Price, & Cook, 2000; but see Müller, Zelazo, Lurye, & Liebermann, 2008). Thus, the criterion for goal neglect set forth by Duncan (1995) is met—children know the appropriate response but do not execute it.

Although there are several disparate accounts of perseveration in the DCCS, they all (explicitly or implicitly) share the central feature that children who perseverate are not holding the correct rule in mind. In the revised version of cognitive complexity and control theory, Zelazo et al. (2003) argue that increases in levels of reflection allow for the formation and implementation of higher order rules that are needed to handle the two contradistinctive rules simultaneously. Children who do not formulate, and thus cannot hold in mind, higher order rules have difficulty on the DCCS. In contrast, children who succeed on the task not only know the rules, but they know that they know the rules allowing them to consider the contradistinctive rules in parallel. According to a different account of perseveration, the attentional inertia account (Kirkham et al., 2003), children perseverate by continuing to act on the wrong rule unless they are primed to activate the correct rule by emphasizing the relevant dimension such as through labeling. Thus, success hinges on accessing the appropriate rule as needed. Similarly, in redescription accounts of the DCCS (Kloo & Perner, 2005; Perner & Lang, 2002), young children cannot think about a stimulus as two things simultaneously (i.e., a *red* thing and a balloon) and therefore it is difficult for them to maintain the new rule in mind, which requires redescribing the stimulus after the rules have been switched. Finally, in accounts that link perseveration to the degree of conflict (Jordan & Morton, 2008; Morton & Munakata, 2002), children's performance can be improved when the degree of conflict is degraded by increasing the salience of the postswitch dimension and the likelihood that the rule is kept in mind. Thus, all the primary accounts of the DCCS attribute perseveration to a failure in accessing the correct rule, which is a central tenet of goal neglect.

Towse, Lewis, and Knowles (2007) investigated goal neglect and its potential causes in young children by combining elements of the DCCS procedure with those of Duncan's (1995) method of inducing goal neglect in adults. Here, 4-year-old children were presented with the outlines of a blue house (on the left) and a red house (on the right) on a laptop computer, and were explained that food items would appear in the houses, but that they should only focus on one particular house. They were taught the rules of the game: (a) On the first presentation of the food items, an arrow will point to the target house that children should initially attend and name the food items. (b) At some unpredictable time, a new task signal will appear: If it is a blue square, then children should begin to name the items in the blue house, but if it is a red square, they should begin to name the items in the red house. Note that these signals sometimes cued the children to continue with the same house, and sometimes required them to switch houses. Because children were tested on whether they understood the rules prior to the study, failure to switch houses after the presentation of the appropriate cue was considered goal neglect. Interestingly, this type of goal neglect correlated negatively with performance on the DCCS, suggesting that goal maintenance is directly related to task switching.

The DCCS has also been modified to assess goal neglect in 4- and 5-year-olds, who typically perform well on the DCCS (Marcovitch, Boseovski, & Knapp, 2007). In the goal-neglect version of the DCCS (GN–DCCS), children received five preswitch trials where they were instructed to sort according to one dimension (e.g., color). On the postswitch trials, children received 30 trials where they were instructed to sort according to the other dimension (e.g., shape). Two types of stimulus cards were used in the postswitch: *conflict* cards, which were the standard DCCS cards where one dimension mapped on to one target card while the other dimension mapped onto the other target card, and *redundant* cards, which were identical to the target cards (see Figure 1). In the *mostly conflict* condition, 80% of the postswitch trials were conflict cards. In contrast, in the *mostly redundant* condition, only 20% of the postswitch trials were conflict cards (i.e., the preponderance of cards were redundant). Marcovitch et al. (2007) reasoned that in the context of constantly sorting conflict cards, it would be relatively simple to keep the relevant postswitch rule in mind either because the conflict trials reinforce the sorting rule, or because the presence of repeated conflict reinforces the need to keep the rule in mind. On the other hand, in the context of constantly sorting redundant cards—where theoretically no rule is needed most of the time, as the test card matches one of the targets on both dimensions—it is more challenging to recall the correct rule when finally faced with a conflict card. Thus,

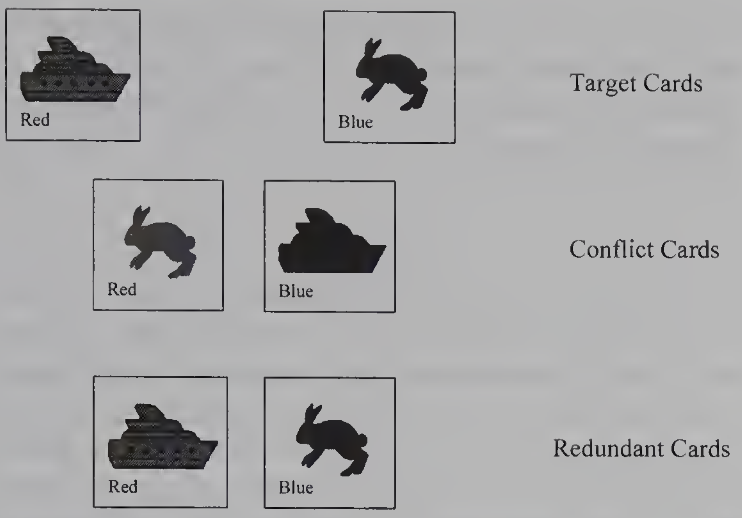


Figure 1. Sample target, conflict, and redundant cards used in the study.
Source: Reproduced from Marcovitch et al. (2007).

Marcovitch et al. predicted that a higher percentage of conflict cards will be sorted correctly in the mostly conflict condition as compared with the mostly redundant condition, which they considered the measure of goal neglect. This is precisely what was found, regardless of whether the children were reminded of the rule once (Experiment 1) or before every postswitch trial (Experiment 2). The latter provides a striking example of goal neglect given that children were told the correct rule immediately prior to sorting each and every card.

In their Stroop study with adults, Kane and Engle (2003) found that performance in a mostly redundant condition (i.e., with most color words presented in their matching hues) was related to individual differences in working memory capacity (WMC). From Kane and Engle's perspective (see also Engle, 2002; Engle & Kane, 2004), WMC reflects "the interaction of attentional and memorial processes in the working memory system, and . . . this interaction between attention and memory is an elementary determinant of broad cognitive ability" (Engle & Kane, 2004, p. 147). Furthermore, they contend that individual differences in WMC are attributed primarily to differences in the control of actively maintaining goal-relevant information. Thus, individuals with high WMC are more likely to maintain a goal actively in the face of distraction, even across extended periods of time where goal activation is only occasionally (and unpredictably) necessary.

Following this perspective, Marcovitch et al. (2007) argued that, compared with adults, 4- and 5-year-old children have difficulty in maintaining goals and that this was the predominant factor for their poor performance on the mostly redundant sort. However, Marcovitch et al. were interested in group performance and did not assess WMC in the children. In the current study, we assessed individual differences in WMC in 4- to 6-year-old children to determine whether WMC was related to goal neglect as measured by the mostly redundant sort. We assessed WMC via three working memory span tasks that all require keeping in mind a series of items while intermittently performing an unrelated, secondary task (e.g., Case, Kurland, & Goldberg, 1982). Given the importance of maintaining representations over time and in the face of distraction on these tasks, we hypothesized that if the mostly redundant card sort indeed relied on these skills then performance should be predicted by individual differences in WMC. Alternatively, it is entirely possible, and consistent with Marcovitch et al.'s (2007) earlier findings, that young children as a whole may not yet have the necessary capacity to maintain goals across periods of inactivity. In other words, even if children's WMC is high relative to their peers, the absolute level of WMC may still not be sufficient to succeed on the mostly redundant sort.

Establishing that individual differences in WMC impact performance on the mostly redundant sort is theoretically and practically important, for several reasons:

1. It would provide a unified explanation for goal neglect from early childhood to adulthood. We hypothesize that goal neglect is a byproduct of the inability to maintain a representation over time and distraction, a skill measured directly by working memory span tasks.
2. A positive relation would imply that improvements in WMC—perhaps via special training—would also improve goal maintenance (e.g., Klingberg et al., 2005). This may be extremely useful for improving goal-directed behavior in typical populations in educational settings, as well as atypical populations in directed training sessions.

We further hypothesize that WMC should be related to performance on the mostly conflict sort for the younger but not the older children. This is because goal maintenance is still required to pass the mostly conflict sort, but the constant usage of the rule makes it easier to keep in mind and partially scaffold performance for the younger children with lower WMC. Older children will generally not require the additional scaffolding to succeed on the mostly conflict sort, and thus no relation between WMC and performance is expected.

Finally, we wanted to assess to what degree WMC predicts performance on each card sort above and beyond the other card sort. On one hand, we expect a meaningful amount of variance common to both card sorts to be accounted for by WMC, as both card sorts have many procedural similarities, and both require goal maintenance albeit to different degrees. However, consistent with the two-factor theory of executive control postulated by Engle and Kane (2004; for a related view, see Braver, Gray, & Burgess, 2007), we also predict that WMC may account for unique variability in the performance of each card sort. According to Engle and Kane, one factor of executive control is "the maintenance of the task goals in active memory" (p. 186), potentially assessed in the current study by performance on the mostly redundant card sort. Engle and Kane argue that goal maintenance is resource-demanding, and thus better executed by individuals with high WMC. The other factor of control is "the resolution of response competition or conflict, particularly when prepotent or habitual behaviors conflict with behaviors appropriate to the current task goal" (p. 186), potentially assessed in the current study by performance on the mostly conflict sort. If each card sort is a measure as a distinct factor of WMC, then variability in performance for each sort should be accounted for uniquely by our measures of working memory.

Method

Participants

We recruited children in a midsized Southeastern city from local day care centers and from a database of parents who had previously expressed interest in participating in research studies. There was no payment for participation, although the children received a gift after the study. The treatment of participants was in accordance with the ethical standards of the American Psychological Association and the study was approved by the local Institutional Review Board. The final sample consisted of thirty-seven 4-year-olds (*M* = 4.71 years, range = 4.02–5.45, 16 girls) and twenty-eight 6-year-olds (*M* = 6.22 years, range = 5.55–6.98, 15 girls). Thirteen additional children (ten 4-year-olds and three 6-year-olds) were tested but their data were not included because they did not perform to criterion on the preswitch sorts (see the following). The sample was racially diverse (69% self-reported as White, 18% Black, 3% biracial, 3% Asian, and 8% preferred not to respond) with varying levels of socioeconomic status (51% reported household earnings over $60,000, 26% under $60,000, and 23% refused to respond).

Materials

For the GN–DCCS, laminated cards (13 cm × 8 cm) depicted either a yellow car, a yellow flower, a green car, a green flower, a red boat, a red bunny, a blue boat, or a blue bunny (all pictures were approximately 16 cm × 16 cm). Participants sorted the cards into two gray sorting boxes (28.5 cm × 20 cm × 11 cm) with slits that had been cut into the lids.

In the Visual Counting Span task of WMC, two types of stickers were affixed to paper cards (21 cm × 20 cm): hearts and American flags, stars and rainbows, or bees and frogs. The number of stickers of each type per card ranged from one to eight. In the Visual Backward Word Span task, paper cards (14 cm × 14 cm) depicted a picture of a common object or animal (approximately 5 cm × 4 cm). Finally, in the Auditory Backward Digit Span task, we presented stimuli using a stuffed pink and white rabbit toy named "Fluffy."

Design and Procedure

In the experimental session, all children completed the three WMC tasks (Visual Counting Span, Visual Backward Word Span, and Auditory Backward Digit Span) and the two card sorting tasks (mostly redundant, mostly conflict). The order of the tasks was counterbalanced with the constraint that the first, third, and fifth tasks were WMC tasks whereas the second and fourth tasks were card sorting tasks.

Goal-Neglect DCCS

All children participated in two games: one *mostly redundant* card sort and one *mostly conflict* card sort (Marcovitch et al., 2007). Children sorted red and blue bunnies and boats in one game and yellow and green cars and flowers in the other. At the beginning of each game, the experimenter affixed the target cards—which differed by both color and shape (e.g., red boat and blue bunny; see Figure 1)—to the sorting boxes. In the five preswitch trials, children were instructed to sort test cards according to one dimension (e.g., color) that was counterbalanced across task order: "In the ________ (shape/color) game, _______ (bunnies/red ones) go here [pointing to the box with the red bunny] and ________ (boats/blue ones) go here [pointing to the box with the blue boat].'' Note that for preswitch trials, all test cards were conflict cards (i.e., the same color as one target, but the same shape as the other target) and the instructions were repeated before every trial. As it is critical for the logic of the task that children pass the preswitch phase, only children who sorted correctly on at least 8 of 10 preswitch trials across both tasks were included in the final sample (see Marcovitch et al., 2007).

On the postswitch trials, the instructions were changed so that children were now required to sort by the other dimension (e.g., shape). Children then received 30 postswitch trials such that the rule was repeated before every fifth trial, regardless of whether it was a redundant or conflict trial. When playing the mostly conflict game, the first 6 trials, and 18 of the remaining 24 trials involved conflict cards (i.e., 80% of all postswitch trials used conflict cards). The remaining 6 trials used redundant cards which matched one of the targets exactly. When playing the mostly redundant game, the contingency was reversed (i.e., 80% of all postswitch trials used redundant cards). In both sorts, the placement of the conflict and redundant cards were randomized for each participant.

Working Memory Tasks

Visual Counting Span (Case et al., 1982). In this task, children were shown cards, one at a time, which displayed pictures of two objects (e.g., frogs and balloons). They were instructed to count one of the objects (e.g., balloons) and remember the number before turning the card face down, after which they were required to recall the numbers in the correct order. Children received three sets each of 2 cards, 3 cards, 4 cards, and 5 cards, in that order.

Visual Backward Word Span (Carlson, Moses, & Breton, 2002). In this task, children were shown cards, one at a time, each of which displayed a single picture familiar to young children (e.g., car, tree, and shoe). The children were asked to identify each picture aloud before turning the card over and then to recall the items in reverse order. Children received three sets each of 2 cards, 3 cards, 4 cards, and 5 cards, in that order.

Auditory Backward Digit Span (Carlson et al., 2002). Children were shown and told about Fluffy, the silly bunny puppet who says things backward. After speaking a list of numbers, the experimenter asked, ''Fluffy says things backwards, so what would he say?'' Children heard three sets each of 2 numbers, 3 numbers, and 4 numbers, in that order.

Results

For all analyses, α is set to .05 for statistical significance and .10 for marginal significance.

Goal-Neglect DCCS

The proportion of conflict cards sorted correctly in each sort is displayed in Figure 2. A 2 × 2 (AGE GROUP × SORT TYPE) mixed analysis of variance (ANOVA) with SORT TYPE as the repeated measure revealed an effect of SORT TYPE, $F(1, 63) = 7.1$, $p < .01$, $MSE = 0.051$, $\eta_p^2 = .10$, such that children sorted more conflict cards correctly in the mostly conflict condition. This replicated the previous finding of goal neglect (Marcovitch et al., 2007). The ANOVA did not reveal an effect of AGE GROUP, $F(1, 63) = 2.0$, $p > .10$, $MSE = 0.11$, $\eta_p^2 = .03$, nor an interaction between AGE GROUP and SORT TYPE, $F < 1$, $MSE = 0.051$, $\eta_p^2 = .00019$.

We also analyzed whether the presentation order of the card sort (ORDER) had an effect on performance, and whether it interacted with AGE GROUP and SORT TYPE. We did find an effect of ORDER, $F(1, 61) = 8.2$, $p < .01$, $MSE = 0.097$, $\eta_p^2 = .12$, such that children who received the mostly redundant card sort first performed better on both card sorts. Notably, there was no interaction between ORDER and AGE GROUP, $F(1, 61) = 1.9$, $p > .10$, $MSE = 0.097$, $\eta_p^2 = .03$, nor between ORDER and SORT TYPE, $F < 1$, $MSE = 0.052$, $\eta_p^2 = .00$, which demonstrates that the effect of goal neglect holds regardless of the order of presentation.

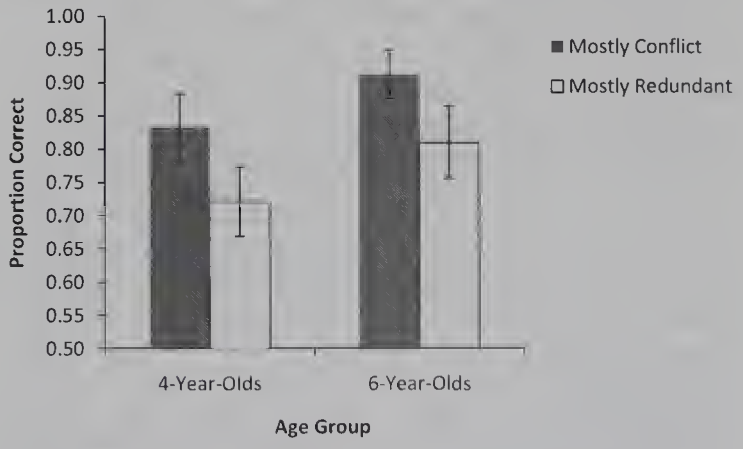


Figure 2. Proportion of conflict cards sorted correctly by age and condition.

Table 1
Means (and Standard Errors) on Working Memory Tasks

	Age group	
Task	4-year-olds	6-year-olds
Visual Counting Span	.59 (.03)	.71 (.03)
Visual Backward Word Span	.63 (.03)	.74 (.02)
Auditory Backward Digit Span	.64 (.04)	.83 (.02)

Working Memory Tasks

For all working memory tasks, each set of items were scored for the proportion of remembered items irrespective of serial order. For example, if the correct response was 2, 3, 6 and the child's response was 2, 6, 4, then that set would be assigned a score of .67 (two correct items out of a possible three items). Then, the average proportion across every set of items was calculated (see Friedman & Miyake, 2005, for advantages using this scoring system).

Means and Correlations Across Tasks

Table 1 displays the means (and standard errors) for all three working memory tasks. As expected, 6-year-olds outperformed 4-year-olds on the Visual Counting Span, $t(63) = 2.7$, $p < .01$, $MSE = 0.028$, $\eta^2 = .10$; the Visual Backward Word Span, $t(63) = 2.9$, $p < .01$, $MSE = 0.023$, $\eta^2 = .11$; and the Auditory Backward Digit Span, $t(63) = 3.5$, $p < .01$, $MSE = 0.048$, $\eta^2 = .16$. Table 2 displays the correlation matrix of all the study variables. To best measure WMC as an individual-differences construct, we created a composite WMC score by standardizing each working memory measure across both age groups and summing them together. Our logic was that because no task is a pure measure of its intended construct, the influence of task-specific variance that is independent of WMC can be minimized, and the influence of WMC-related variance maximized, by averaging across multiple WMC tasks that each make different non-WMC-related demands (see, e.g., Conway et al., 2005; Kane et al., 2004).

WMC as a Predictor of the GN–DCCS

Separate regression analyses were conducted on the proportion of correctly sorted conflict cards in the mostly redundant sort (PROP REDUNDANT), our measure of goal maintenance, and the proportion of correctly sorted conflict cards in the mostly conflict sort (PROP CONFLICT), our measure of response competition. In these analyses, the predictor variables were age (AGE, standardized), the WMC composite (standardized), and the interaction between the two predictor variables (AGE × WMC). Given the multicollinearity between AGE and WMC ($r = .57$, $p < .01$), we opted to enter WMC, the variable of primary importance, into the model as our first step. Then, we entered AGE as our second step and AGE × WMC as our last step.

The regression analysis on PROP REDUNDANT revealed a good overall fit, $F(3, 61) = 2.9$, $p < .05$, $R^2 = .13$, and a significant effect of WMC, $B = .11$, $t(61) = 2.8$, $p < .01$, $\Delta R^2 = .11$. Neither AGE ($\Delta R^2 = .01$) nor AGE × WMC ($\Delta R^2 = .002$) added significantly to the model. Similarly, the regression analysis on PROP CONFLICT revealed a good overall fit, $F(3, 61) = 3.8$, $p < .05$, $R^2 = .16$, and a significant effect of WMC, $B = .10$, $t(61) = 2.1$, $p < .05$, $\Delta R^2 = .15$. Once again, neither AGE ($\Delta R^2 = .005$) nor AGE × WMC ($\Delta R^2 = .004$) added significantly to the model. Thus, our predictions were confirmed that WMC is a significant predictor of performance on both the mostly redundant and mostly conflict card sorts.

Table 2
Correlation Matrix of Study Variables

	Card sorts		Working Memory Span tasks		
	Mostly redundant	Mostly conflict	Visual Counting	Visual Word	Auditory Digit
Age	.28*	.16	.41**	.41**	.55**
Mostly redundant		.39**	.24†	.25*	.31*
Mostly conflict			.38**	.23†	.31*
Visual Counting				.59**	.35**
Visual Word					.47**

†$p < .10$. *$p < .05$. **$p < .01$.

It is also possible that the multicollinearity of the predictors suppressed age differences in our analyses. To determine whether there were differences between age groups, we conducted additional regression analyses for each age group with WMC as the predictor variable. For the mostly redundant card sort, the regression analysis with the 4-year-olds revealed a marginally significant effect of WMC, $B = .03$, $t(35) = 1.7$, $p < .10$, $R^2 = .08$, but a significant effect of WMC with the 6-year-olds, $B = .07$, $t(26) = 2.1$, $p < .05$, $R^2 = .14$, confirming that WMC predicts performance across both ages. In contrast and as expected, the regression analysis for the mostly conflict card sort revealed a significant effect of WMC for the 4-year-olds, $B = .05$, $t(35) = 2.8$, $p < .01$, $R^2 = .18$, but not the 6-year-olds, $B = .01$, $t < 1$, $R^2 = .01$.

Finally, we sought to determine whether WMC predicted variance for each sort above and beyond the other sort. To accomplish this, we included PROP CONFLICT in the first step of the regression on the mostly redundant sort, and likewise included PROP REDUNDANT in the first step of the regression on the mostly conflict sort. For these analyses we once again entered AGE as a predictor (i.e., we analyzed the entire sample together).

The regression analysis for the mostly redundant card sort revealed that PROP CONFLICT was a significant predictor of performance, $\Delta R^2 = .15$, $p < .01$, and that WMC accounted for a marginally significant amount of additional variability, $\Delta R^2 = .04$, $p < .10$. Similarly, the regression analysis for the mostly conflict card sort revealed that PROP REDUNDANT was a significant predictor of performance, $\Delta R^2 = .15$, $p < .01$, and that WMC accounted for a significant amount of additional variability, $\Delta R^2 = .07$, $p < .05$. In neither case did AGE nor AGE × WMC contribute significantly to the model.

Discussion

WMC has been associated with a wide range of cognitive skills that involve shifts of attention and executive control (Engle & Kane, 2004; Gathercole et al., 2008). The current study was designed to assess whether WMC was predictive of the ability to maintain a novel goal in young children. The results demonstrated that both 4- and 6-year-old children with higher WMC were more successful in goal maintenance, as measured by performance on the mostly redundant sort. This finding is consistent with Kane and Engle's (2003) report that WMC was related to goal maintenance in adults and extends previous research by demonstrating the relation between working memory and goal maintenance in young children. The current findings refute the idea that young children in general simply do not have the capacity to maintain a representation over a period of inactivity, and thus emphasize the importance of individual differences in WMC on the execution of goal-directed behavior.

WMC also predicted performance on the mostly conflict sort for 4-year-olds but not for 6-year-olds. We interpret 4-year-olds' higher performance on the mostly conflict sort compared with the mostly redundant sort as evidence that they benefit from the constant execution of the rule. However, because they still have some difficulty maintaining the rule, there are individual differences based on WMC. In contrast, 6-year-olds generally possess a higher WMC (relative to 4-year-olds) that is sufficient to take full advantage of the scaffolding afforded by the mostly conflict task, and so WMC-related individual differences were absent. In addition, the success of 6-year-olds (relative to 4-year-olds) on the mostly conflict sort must be independent of WMC and suggests that other age-related processes might be contributing to success. One candidate is the ability to switch flexibly from one response set to another (i.e., set shifting; Cragg & Chevalier, in press; Garon, Bryson, & Smith, 2008; Miyake et al., 2000), a process that is not tested directly in our measure of WMC but is clearly needed to succeed on the mostly conflict sort. Indeed, vast improvements in set shifting have been reported in young children, especially between 3 and 6 years of age (e.g., Chevalier & Blaye, 2008; Deak, 2000; Zelazo et al., 2003).

We argue that the two card sorts are tapping into related but separable abilities. This is because: (a) WMC predicts common variance for both card sorts along with unique variance for each card sort, (b) the modest correlation between performance on these two sorts ($r = 0.39$, $p < .01$) suggests that these factors are interrelated but separable, and (c) there are different patterns of performance for 4- and 6-year-olds on the card sorting tasks. As argued earlier, this is consistent with Engle and Kane's (2004) two-factor theory of working memory and extends the theory to early childhood. Importantly, we now have some empirical evidence that each factor can be measured by its own card sort. Clearly, additional research needs to be conducted that speaks to the extent that each card sort assesses each factor, and the importance of assessing the development of each factor independently.

The findings from the current study have direct implications for implementing ways to improve goal-directed behavior. From our perspective, goal maintenance can be enhanced in children who undergo specific training in working memory. This idea is supported by recent research from Diamond, Barnett, Thomas, and Munro (2007) who have used the Tools curriculum to train low income preschool children on a variety of executive function tasks in an effort to improve executive control, the central component of working memory as conceptualized by Engle and Kane (2004; see also Baddeley, 2007). Vast improvements were seen on several measures, including the Dots task, which interleaves congruent and incongruent trials in a similar manner to the GN–DCCS.

In sum, individual differences in working memory predict goal neglect even in young children. These results highlight the utility of this paradigm to set the groundwork for future studies designed to investigate the specific mechanisms involved in the development of working memory.

References

Baddeley, A. D. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*, 111–127.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). New York: Oxford University Press.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory *Infant and Child Development, 11*, 73–92.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science, 11*, 339–353.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769–786.

Cragg, L., & Chevalier, N. (in press). The processes underlying flexibility in childhood. *Quarterly Journal of Experimental Psychology*.

Deak, G. O. (2000). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior, 31*, 271–327.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*, 257–303.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York: Academic Press.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*, 581–590.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliot, J. G., Holmes, J., & Hilton, K. A. (2008). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences, 18*, 214–223.

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology, 35*, 770–780.

Jordan, P. L., & Morton, J. B. (2008). Flankers facilitate 3-year-olds' performance in a card-sorting task. *Developmental Psychology, 44*, 265–274.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47–70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The domain generality of working-memory capacity: A latent-variable approach to verbal and spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kirkham, N. Z., Cruess, L. M., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–467.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., et al. (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. *Journal of the*

American Academy of Child and Adolescent Psychiatry, *44*, 177–186.

Kloo, D., & Perner, J. (2005). Disentangling dimensions in the Dimensional Change Card Sorting task. *Developmental Science*, *8*, 44–56.

Marcovitch, S., Boseovski, J. J., & Knapp, R. J. (2007). Use it or lose it: Examining preschoolers' difficulty in maintaining and executing a goal. *Developmental Science*, *10*, 559–564.

Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, *12*, 1–18.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*, 947–959.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ''frontal lobe'' tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage in early childhood. *Developmental Psychobiology*, *40*, 255–265.

Müller, U., Zelazo, P. D., Lurye, L. E., & Liebermann, D. P. (2008). The effect of labeling on preschool children's performance in the Dimensional Change Card Sort Task. *Cognitive Development*, *23*, 395–408.

Perner, J., & Lang, B. (2002). What causes 3-year olds' difficulty on the Dimensional Change Card Sorting task? *Infant and Child Development*, *11*, 93–105.

Towse, J. N., Lewis, C., & Knowles, M. (2007). When knowledge is not enough: The phenomenon of goal neglect in preschool children. *Journal of Experimental Child Psychology*, *96*, 320–332.

Towse, J. N., Redbond, J., Houston-Price, C. M. T., & Cook, S. (2000). Understanding the Dimensional Change Card Sort: Perspectives from task success and failure. *Cognitive Development*, *15*, 347–365.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, *68* (Serial No. 274).

Child Development, November/December 2010, Volume 81, Number 6, Pages 1696–1713

The Social Costs of Academic Success Across Ethnic Groups

Thomas E. Fuller-Rowell
University of Michigan

Stacey N. Doan
Boston University

This study explores the longitudinal association between academic achievement and social acceptance across ethnic groups in a nationally representative sample of adolescents (N = 13,570; M_{age} = 15.5 years). The effects of school context are also considered. Results show that African American and Native American adolescents experience greater social costs with academic success than Whites. Pertaining to school context, findings suggest that the differential social consequences of achievement experienced by African Americans are greatest in more highly achieving schools, but only when these schools have a smaller percentage of Black students. Students from Mexican descent also showed differential social costs with achievement in particular contexts. The implications of these findings to theory, policy, and future research are discussed.

Despite significant progress being made toward closing ethnic gaps in achievement, their relative stability over the past two decades has raised their priority within the broader political agenda and caused them to become recognized as one of the most important civil rights issues of the 21st century (Paige, 2004). Across virtually all measures, Black, Hispanic, and Native American students in the United States earn lower grades, drop out more often, and attain less education than Whites or Asian Americans (Perie & Moran, 2005). While structural and social burdens, often experienced disproportionately by minorities, such as socioeconomic status (McLoyd, 1990), single-parent families (Pong, 1998), and school or neighborhood disadvantage (Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993) are important factors to consider, the dynamics of social acceptance among academically stigmatized groups are also of theoretical and practical interest (Ogbu, 2004; Spencer, Cross, Harpalani, & Goss, 2003; Steele, 1998).

Social acceptance, particularly critical during adolescence when the opinions and judgments of peers become increasingly important, is considered a basic need, closely tied to motivation and behaviors (Baumeister & Leary, 1995). Research exploring the association between achievement and social acceptance suggests a positive cyclical relation, achievement leading to greater social acceptance and social acceptance leading to greater levels of achievement (Chen, Rubin, & Li, 1997; Wentzel, 1991, 2005). These findings, combined with evidence that social acceptance is an essential component of adolescent self-worth (Harter, 1999), suggest that any breakdown in this relation could have particularly adverse consequences to development. The extent to which particular ethnic groups experience differential social costs with achievement is therefore an important area of research and theory.

Theories Predicting Differences Across Ethnic Groups in the Social Costs of Achievement

To understand the current experiences and perspectives of a particular minority group, it is important to consider the history of the group's relationship to the dominant culture (Fordham & Ogbu, 1986; Ogbu, 1999, 2004). Specifically, in their oft-cited paper, Fordham and Ogbu (1986) argue that involuntary minority groups, such as African Americans and Native Americans, whose presence

---

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from Grant P01-HD31921 for this analysis.

Correspondence concerning this article should be addressed to Thomas E. Fuller-Rowell, Department of Psychology, Institute for Social Research, University of Michigan, 426 Thompson Street, #5132, Ann Arbor, MI 48106. Electronic mail may be sent to tf42@cornell.edu or tefr@umich.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0007

in the United States stems back to colonization and enslavement, have developed a collective identity in opposition to the mainstream White culture. Furthermore, Ogbu and colleagues suggest that, for these groups, academic success may be stigmatized by peers as ''acting White,'' and thus attainment of higher grades in school may have differential social consequences (Ogbu & Simons, 1998).

Several scholars, however, have challenged the conceptual underpinnings of this perspective. Cross (2003), for example, has argued that resistance to education is in no way fundamental to African American culture and that the roots of achievement problems lie not in the legacy of slavery but within specific structural inequalities that have and still continue to directly affect minority groups. Furthermore, Spencer and Harpalani (2008) argue that Ogbu and colleagues make broad unwarranted assumptions that decontextualize and overgeneralize the experiences of African American youth while ignoring the importance of meaning-making processes. In particular, Spencer and Harpalani emphasize that the ''acting White'' phenomenon is not a cultural syndrome that is pervasive, but rather a coping mechanism: a reaction to stereotypes and experiences of discrimination experienced in particular contexts.

Spencer and colleagues therefore articulate an alternative interpretation of the ''acting White'' phenomenon that focuses on processes of identity development, and experiences with stigma and discrimination (e.g., Spencer & Harpalani, 2008; Spencer, Noll, Stoltzfus, & Harpalani, 2001; Spencer et al., 2003). Important to this perspective are normative developmental processes during adolescence. Specifically, adolescence is an important time for youth to engage in identity exploration in various domains (e.g., career, ethnic identity). Although members of the majority group generally take their ethnic identity for granted, ethnic minority adolescents have the additional task of having to negotiate the meaning of their membership in a group whose customs may often be devalued by mainstream society (Cross, 1991; Phinney, 1990; Spencer & Markstrom-Adams, 1990). As part of this identity formation process, Cross (1991) has described that ethnic minorities often go through a stage referred to as immersion-emersion, in which they immerse themselves within their own ethnic culture and have a tendency to reject the perspectives of the dominant group.

During this sensitive stage of identity development, minority adolescents tend to react to negative academic stereotypes directed toward their group in a particular context by labeling behaviors associated with success within that context as ''acting White'' (Spencer et al., 2003). Spencer and colleagues, therefore suggest that the ''acting White'' phenomenon is not the result of a broad cultural frame of reference (as Ogbu's theory suggests) but rather the result of a particular coping response to negative stereotypes that are experienced in specific contexts. From this perspective, it is the current stigmatization of minority groups that drives some adolescents to reject expectations and values of the dominant group, rather than any inherent oppositional orientation toward schooling. Under Spencer's framework, stigmatized minority groups other than African Americans would therefore also be expected to experience social costs with academic success. Furthermore, contextual factors, such as school characteristics, would also be expected to play an important role.

The theory of stereotype threat is also of direct relevance to the proposition that academic achievement may be coming at a greater social cost for particular groups. Steele (1997, 1998) has discussed how the burden of academic stereotypes, and the prolonged exposure to such stereotypes at the group level, may create social costs with achievement for members of stigmatized groups. According to this perspective, when individuals become aware of a negative stereotype directed toward their group, they experience anxiety related to the possibility of conforming to the stereotype, which in turn affects their performance. Thus, as a self-protective mechanism, prolonged exposure to negative stereotypes in a particular domain is thought to be associated with psychological disengagement from that domain (Davies, Spencer, & Steele, 2005; Osborne, 1997; Steele, 1997). Steele's work, therefore, suggests that the collective devaluing of academics among stereotyped groups may lead individuals who are part of those groups to experience greater social costs with achievement (Steele, 1997).

Related Empirical Findings

Support for the ''acting White'' proposition has been derived from several ethnographic studies of African American students, conducted across multiple contexts (e.g., Fordham & Ogbu, 1986; Ogbu, 1999). These studies generally found that attitudes and behaviors conducive to academic attainment, such as studying, reading, and participating in class, are often stigmatized as ''selling out'' or ''acting White'' (e.g., Fordham & Ogbu, 1986).

Other more recent qualitative studies, however, have failed to find that Black adolescents experience any oppositional orientation toward achievement (e.g., Akom, 2003; Tyson, 2002).

Additionally, the few quantitative studies that have examined the social costs of academic success have thus far failed to establish that achievement leads to greater social costs for particular groups. Specifically, using data from the National Educational Longitudinal Study (NELS), two studies have attempted to test the proposition that African Americans experience greater social costs with academic success (Ainsworth-Darnell & Downey, 1998; Cook & Ludwig, 1997). The NELS survey included self-report measures of academic achievement, popularity, and harassment. Using these measures, both studies found no evidence to suggest that African American students experienced greater social costs with achievement.

While this research calls into question the social costs proposition, findings from these studies are limited in several ways. First, both studies are based on the same dataset and therefore do not offer an independent replication of findings. Second, neither study uses a longitudinal design to test the association between achievement and social acceptance: Both studies only considered associations between self-report items within a single time point. Furthermore, these studies only look at a narrow age range and do not compare social costs across all of the major U.S. pan-ethnic groups. Finally, neither study uses multilevel modeling techniques, which allow for a more adequate consideration of individual and school-level variables (Raudenbush & Bryk, 2002).

The Importance of School Context

Individuals develop within particular environmental contexts, and these contexts play a pivotal role in development (Bronfenbrenner, 1979; Spencer, 1999). The environment of a school, specifically its achievement level and the proportion of same-race students in the school, may play an important role in moderating the relation between achievement and social acceptance. Research examining intergroup relations, for example, has demonstrated that in competitive or high-achieving contexts, where both groups are being evaluated on the same criterion, racial tensions and discrimination are likely to increase (see Brewer & Kramer, 1985 for a review). Furthermore, recent qualitative data suggest that competitive schools may breed an environment of animosity, particularly when there is a disproportionate underrepresentation of disadvantaged students (Tyson, Darity, & Castellino, 2005).

The proportion of same-race students in a school is therefore also an important factor that may influence the social costs of academic success for minority adolescents. Specifically, with more same-race students present, it is possible that experiences of stigmatization and discrimination may occur less often or be less pronounced. Along these lines, researchers have argued that in largely Black schools, students are less likely to associate racial characteristics with achievement and therefore less likely to stigmatize achievement-related behaviors (O'Connor, Fernandez, & Girard, 2007). Furthermore, students in largely Black schools have been found to hold more proschool attitudes (Goldsmith, 2004), and higher levels of school attachment (Johnson, Crosnoe, & Elder, 2001). This work suggests that the social costs of academic success may be ameliorated for minority students in schools with a higher proportion of same-race students.

Purposes of the Current Study

The purpose of this article is to explore the association between achievement and social acceptance across ethnic groups in a nationally representative sample of adolescents. Specifically, we will consider the social costs of achievement for all of the major pan-ethnic groups in the United States (non-Hispanic Whites, African Americans, Hispanics, Asians, and Native Americans), as well as subgroups within the Hispanic and Asian pan-ethnic categories. We will then determine whether socioeconomic and contextual factors (family socioeconomic status [SES], school SES, family structure, school-level achievement, school safety, school type, and school size) are able to account for any differences in social costs across groups. Based on the theories discussed (Ogbu & Simons, 1998; Spencer & Harpalani, 2008; Steele, 1998), we hypothesize that even after accounting for background factors, stigmatized ethnic groups, such as African Americans, Hispanic Americans, and Native Americans, will, on average, experience greater social costs with achievement than Whites.

Having established differences across groups, we will then explore factors that may account for variation in social costs within groups. Specifically, gender and immigrant status will be considered as potential moderators of social costs. With respect to gender, previous work suggests that the dynamics of stigma and achievement are somewhat more problematic for African American men than women (e.g., Graham, Taylor, & Hudley, 1998; Osborne, 1997). Thus, gender will be explored as a potential moderator.

Additionally, a number of scholars have argued that immigrant status is an important factor explaining differences in attitudes toward academic attainment (e.g., Portes & Zhou, 1993; Spencer et al., 2003). Furthermore, research has indicated that immigrant status influences the extent to which discrimination is perceived (Finch, Kolody, & Vega, 2000). Based on this work, we will test the hypothesis that more recent Hispanic, Asian, or Black immigrants experience lower social costs with academic success than their more recent immigrant counterparts.

Our final analyses will focus on exploring whether social costs are dependent on school context. Specifically, based on the research discussed before, we will consider whether the racial composition and achievement level of a school moderate the relation between academic achievement and social acceptance. In relation to school context, we hypothesize that minority groups will show higher social costs with academic success in higher achieving schools. We also hypothesize that the presence of same-race students will serve as a protective factor in these contexts, buffering the level of social costs experienced with academic achievement.

Method

Data

The current study utilizes data from the National Longitudinal Study of Adolescent Health (Add Health). Add Health is an ongoing nationally representative study of 7th through 12th graders selected from 80 high schools and 52 feeder (middle and junior high) schools in 1994. All available students in each of the participating schools initially completed the In-School Questionnaire ($n = 90{,}118$). A subsample of students then completed the Wave I In-Home Interview ($n = 20{,}745$) in 1995, followed by the Wave II In-Home Interview in 1996 ($n = 14{,}738$). In-Home Interviews can also be linked to school-level data reported by school administrators, as well as to data from parent interviews.

Sample

The sample for the current study consisted of adolescents who participated in both the Wave I and Wave II In-Home Interviews, and were assigned a valid sample weight as part of the Add Health nationally representative sample ($n = 13{,}570$; see Chantala & Tabor, 1999, for details on the Add Health sampling procedures). Sample descriptive statistics by pan-ethnic group are presented in Table 1, and are described in the Results section.

Models containing only individual-level variables (Models 1–3, 5, and 6) excluded 5% of cases owing to missing values on one or more covariates. Additional models (Model 4, and Models 7 through 15), also excluded schools with missing school-level data and therefore had slightly higher levels of overall missingness (approximately 7% of cases). Those who were excluded from analyses as a result of missing data were not found to be different from those included in the analysis on any of the background or substantive variables considered in the

Table 1
Descriptive Statistics by Ethnic Group

	White	Black	Hispanic	Asian	Native	Mixed	Other	$F(6, 123)$	η^2
Age	15.46	15.71	15.63	15.77	15.38	15.38	15.22	1.36	.005
Male (%)	50.0	49.2	51.1	53.8	64.0	49.4	52.2	1.31	.001
SES	2.89_a	2.54_b	2.11_c	3.06_a	2.42_b	2.72_b	2.70_b	17.79***	.063
Single-parent family (%)	26.1_b	58.1_a	33.2_b	16.5_c	46.6_a	38.3_b	35.8_b	50.28***	.061
Foreign-born parent (%)	3.6_d	3.3_d	55.9_b	82.2_a	1.2_d	14.7_c	46.6_b	60.55***	.399
GPA (Wave I)	2.87_b	2.58_d	2.61_d	3.19_a	2.49_d	2.70_c	$2.80_{b,c}$	22.74***	.036
Social acceptance (Wave I)	0.051_a	-0.068_b	-0.030_a	-0.116_b	$-0.184_{a,b}$	$-0.071_{a,b}$	$-0.051_{a,b}$	2.81**	.003
Social acceptance (Wave II)	0.070_a	-0.033_b	-0.119_b	-0.126_b	$-0.039_{a,b}$	-0.158_b	$-0.224_{a,b}$	7.20***	.007
n	7,051	2,663	2,132	843	78	688	104		

Note. Table values are population point estimates or proportions which account for the Add Health sampling strategy. Different subscripts across rows represent ethnic group contrasts that are significant at the 95% confidence level. SES = socioeconomic status; GPA = grade point average.
$^{**}p < .01$. $^{***}p < .001$.

study (all point-biserial and phi correlations were below a magnitude of .10). The exclusion of cases with missing data was therefore assumed to add little substantive bias to the reported results.

Measures

Race and ethnicity. Racial and ethnic categories were created based on self-report items from the Wave I In-Home Interview. Participants first reported whether they are of Hispanic origin. Next, participants indicated the racial category or categories they belong to (Asian, Black or African American, Native American, White, or Other). Those who reported being part of two or more categories were classified as mixed racial, and the remaining individuals were classified into the standard categories of non-Hispanic Black, non-Hispanic White, non-Hispanic Asian, non-Hispanic Native American, and Other. Sample sizes and demographics for each racial category are included in Table 1. In addition to the pan-ethnic groups described before, individuals who identified themselves as Hispanic or Asian also identified the specific subgroup(s) that they belong to. For Hispanics, we consider Mexican ($n = 1041$), Puerto Rican ($n = 336$), Cuban ($n = 294$), and Central or South American ($n = 195$) groups separately for some analyses. For Asians, we consider Chinese ($n = 191$) and Filipino ($n = 360$) groups separately. Other Asian subgroups were not considered because of small sample sizes.

Grade point average. Grade averages at Wave I were calculated from self-reports of achievement in each of the four major subject areas (English or language arts, history or social studies, mathematics, and science). Participants were asked to report their grade at the most recent grading period on a 4-point scale from *A* to *D or lower*. The four items were averaged to create an overall score ranging from 1 (*all Ds or lower*) to 4 (*all As*) representing each participant's grade point average (GPA) at Wave I (Cronbach's $\alpha = .75$). The four item scale was then standardized for use in the reported models. Self-reported GPA has been shown to be highly correlated with actual GPA and therefore can be considered an adequate proxy for actual levels of achievement (Gonzales, Cauce, Friedman, & Mason, 1996).

Social acceptance. Social acceptance was measured using four items from the Wave I and Wave II In-Home Interviews. Scale items map closely onto established conceptualizations of perceived social support, loneliness, and sense of belonging (e.g., Hagerty & Patusky, 1995; Russell, 1996). The term *social acceptance* is therefore used broadly in this article to encompass all three of these closely related concepts. The first item asked participants how strongly they agree or disagree that they "feel socially accepted." Responses for this item were on a 5-point scale ranging from *strongly agree* to *strongly disagree.* The remaining three items asked participants to report how often in the past week "people were unfriendly" to them, how often they "felt that people disliked" them, and how often they "felt lonely." Responses to these items were on a 4-point scale ranging from *never or rarely* to *most of the time or all of the time.* Items were standardized, and averaged such that higher scores indicated higher levels of social acceptance (lower levels of social isolation). A log transformation was performed to reduce skewness and robust standard errors were used in all reported models to account for remaining deviations from normality. The average interitem correlation for the scale was .32, and the internal reliability was .65 at Wave I and .66 at Wave II. These are equivalent to reliabilities of .79 and .81 for an eight-item scale with equivalent interitem correlations (Cronbach, 1951). Internal reliability of the scale was also very similar across ethnic groups (.66 for Whites, .64 for African Americans, .65 for Hispanics, .63 for Asians, and .63 for Native Americans). Furthermore, exploratory and confirmatory factor analyses both suggested that a single factor structure was the most appropriate solution for all five pan-ethnic groups.

Individual-level disadvantage. SES and family structure were considered as indicators of individual-level disadvantage. SES was measured from youth reports of the highest level of education obtained by a currently residing parent. Level of education was recoded to the following 4-point scale: 1 = *less than high school,* 2 = *graduated from high school or obtained a GED,* 3 = *some college or post-high school technical training,* and 4 = *graduated from college or more.* Where youth reports were missing, parent self-reports were used to minimize missing data (parents responded to an identical question with the same response categories). Family structure was dummy coded with 1 indicating single parent family.

Other individual-level variables. Age, gender, and immigrant status were also included in multilevel models. Immigrant status and age were based on youth reports. Immigrant status was dichotomously coded with 1 indicating a foreign-born mother or primary caregiver. Coding immigrant status in this manner is synonymous with comparing first- and second-generation adolescents (foreign-born parents [FBPs]) with third-generation or greater adolescents (U.S.-born parents). Age was mean centered so that the model intercepts remained interpretable as the average aged adolescent. Gender was recorded by the Add Health interviewer during the Wave I In-Home Interview and was dichotomously coded with 1 indicating men. Population means and standard deviations for all individual-level variables are reported in Table 2.

School-level variables. At the school level, five measures of school disadvantage were considered: safety, SES, achievement, size, and type (public vs. private or Catholic). Means and standard deviations for all school-level variables are reported in Table 3. With respect to safety, all students who completed the In-School Survey (a near census of each school) were asked how strongly they agree or disagree that they "feel safe at school." By aggregating responses to this item, average levels of safety were calculated for each school. An identical procedure was also carried out to compute an aggregated school-level SES score based on student reports of parent education from the In-School Survey. This aggregation method is akin to techniques used in previous Add Health studies (e.g., Crosnoe, Cavanagh, & Elder, 2003).

School-level achievement was measured using indicators of grades and test scores. Although both grades and test scores have unique limitations as

Table 2

Correlations and Descriptive Statistics for Individual-Level Variables (Level 1)

Variables	1	2	3	4	5	6	7	8
Age	—							
Male	.05	—						
SES	−.06	$.02_{ns}$	—					
Single parent	.05	$-.02_{ns}$	−.23	—				
Foreign-born parent	.05	$.01_{ns}$	−.12	−.04	—			
GPA (Wave I)	−.13	−.14	.27	−.17	.03	—		
Social acceptance (Wave I)	−.11	.09	.09	−.08	$-.02_{ns}$	.13	—	
Social acceptance (Wave II)	−.07	.08	.10	−.08	−.04	.13	.49	—
M	15.52	0.50	2.74	0.32	0.13	2.79	0.01	0.02
SD	1.61	0.50	1.06	0.47	0.34	0.78	1.00	1.00

Note. All table values are population estimates that account for the Add Health sampling strategy. All correlations with a magnitude of .03 or larger are significant at the 95% confidence level. *ns* = not significant; SES = socioeconomic status; GPA = grade point average.

Table 3

Correlations and Descriptive Statistics for School-Level Variables (Level 2)

Variables	1	2	3	4	5	6	7	8	9	10	11
Black school[a]	—										
% Black	.87***	—									
Hispanic school[b]	.01	.00	—								
% Hispanic	.05	.08	.75***	—							
% Asian	−.07	−.05	.42***	.31***	—						
% Native American	−.09	−.13	.03	.19*	−.11	—					
SchSES	−.08	.02	−.24**	−.24**	.17*	−.25**	—				
SchSafety	−.48***	−.40***	−.12	−.25**	−.11	−.16*	.40***	—			
SchAchieve	−.30*	−.30**	−.05	−.28**	.00	−.13	.55***	.57***	—		
Log size	.29*	.24**	.17*	.21*	.25**	−.12	−.08	−.56***	−.27**	—	
Not public[c]	−.13	−.08	.07	.05	.12	−.14	.33***	.36***	.18*	−.16*	—
M	0.20	0.13	0.11	0.09	0.02	0.02	2.84	3.94	0.31	5.38	0.14
SD	0.40	0.23	0.32	0.12	0.04	0.03	0.38	0.34	0.89	1.28	0.35

Note. All table values are population estimates which account for the Add Health sampling strategy. SchSES = school-level socioeconomic status; SchSafety = school-level safety; SchAchieve = school-level achievement.

[a]Black school: 0 = lower three quartiles of percent Black; 1 = top quartile percent Black. [b]Hispanic school: 0 = lower three quartiles of percent Hispanic; 1 = top quartile percent Hispanic. [c]Not public: 0 = public; 1 = private or Catholic.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

measures of school-level achievement (Hanushek & Taylor, 1990; Willingham, Pollack, & Lewis, 2002), the two measures together are better able to capture achievement at the school level (Clark & Watson, 1995). School-level grades were calculated by aggregating student reports of their GPA from the In-School Survey. School test scores, on the other hand, were taken from school administrator reports of the percentage of students testing above and below grade level. The percentage of students testing below grade level was subtracted from the percentage of students testing above grade level, yielding a measure of the extent to which there was a preponderance of students in a given school with high test scores. These two indicators of school-level achievement were then standardized and averaged ($r = .46$, $p < .001$). School size and school type were based on school administrator reports of the size of the school (total student enrollment) and whether the school was public, private, or Catholic. The school size variable was log-transformed to reduce right skewness, and a dichotomous variable was created for school type such that 1 indicated a school that was not public (i.e., private or Catholic).

With respect to the percentage of Black students, a dichotomous school-level variable was created indicating a school in the top quartile with respect to the percentage of Black students. Of the 132 schools surveyed as part of the Add Health nationally representative sample, 33 fell into this category. These schools had student bodies with an average of 47% Black students ($SD = 20.7$). This variable therefore indicates schools where Black students are either the majority ethnic group in the school, or a substantial portion of the overall student body. The use of a dichotomous variable in this context is in line with the perspective that a critical mass of a particular group is necessary to change intergroup dynamics (Etzkowitz, Kernelgor, Neuschatz, Uzzi, & Alonzo, 1994). A second dummy variable was also created to indicate schools in the top quartile with respect to the percentage of Hispanic students. Twenty-four schools fell into this category. These schools had an average of 42% Hispanic students ($SD = 20.6$). While continuous variables relating to the percentage of Black and Hispanic students in a school were also explored, we found that dichotomous variables were better able to capture differences between schools in social costs for particular groups. The indicator variables are therefore used in the analyses presented in this article. All continuous variables were z-scored in the models presented.

Analysis Plan

Analyses for the current study involved multilevel modeling of individual- and school-level variables (Raudenbush & Bryk, 2002). Because Add Health data collection was school-based, we included a random effect for school in each of our models as well as a random slope component for the relation between grades and social acceptance. Including random effects allowed us to account for within-school clustering of the data and therefore to accurately assess the significance of model parameters. Multilevel models also allowed us to consider individual-level variables, school-level variables, and cross-level interactions simultaneously.

Four multilevel models were estimated to address questions relating to ethnic group differences in the relation between achievement and social acceptance: Model 1 focused on establishing the overall effect of GPA at Wave I on subsequent changes in social acceptance, Model 2 tested whether this relation differed across ethnic groups, and Models 3 and 4 determined whether differences across groups could be explained by other individual- and school-level factors. After establishing the relation between grades and social acceptance for each ethnic group, the next model focused on testing whether immigrant status may be affecting social costs for particular ethnic groups. Specifically, Model 5 considered whether Hispanic, Asian, or Black youth from immigrant families show lower social costs with achievement than those from families with U.S.-born parents. Model 6 then considered whether African American men experienced more social costs than women.

The last set of models focused on exploring potential school-level contingencies of the relation between grades and social acceptance: Model 7 looked at whether social costs for Black students were greater in schools with higher levels of achievement, Model 8 looked at whether Black students in largely Black schools experienced lower social costs than those in less Black schools, and, finally, Model 9 considered whether the effects of school achievement on social costs depended on the proportion of Black students in the school.

Similar models also tested the social consequences of achievement for Hispanic students as a function of school-level achievement and the percentage of Hispanic students in the school (Models 10, 11, and 12). The effects of school context were also considered for individual Hispanic ethnic groups (Mexican, Puerto Rican, Cuban, and Central or South American) and Asian ethnic groups

(Chinese, and Filipino). Models 13, 14, and 15, for example, tested the effects of school contexts for adolescents from Mexican descent, and subsequent models tested context effects for the other Hispanic and Asian ethnic groups. Context effects could not be considered for Native Americans because of sample size limitations.

Because the Add Health sampling frame was stratified by various school characteristics, and because various subgroups of individuals were oversampled, it was important to account for sample weights in all of our analyses. In addition to adjusting for the probability of selection at the level of both school and individual, sample weights also adjust for survey nonresponse between the Wave I and Wave II In-Home Interviews (Chantala & Tabor, 1999). This corrects for potential bias to the sample that may have been added with attrition between the first and second waves. All of the models presented in the current study are estimated using the gllamm procedure within Stata (Rabe-Hesketh, Pickles, & Skrondal, 2001; StataCorp, 2007).

Results

Descriptive Analysis and Correlations

All analyses presented in this article account for the Add Health sampling strategy and are therefore representative of the adolescent population in the United States at the time the data were collected. Means differences were first considered across the different ethnic groups. Overall, ethnicity accounted for a significant portion of the variance in each of the variables considered. Ethnic group differences among individual-level variables are presented in Table 1.

Correlations between individual-level variables are presented in Table 2 along with estimated population means and standard deviations for each variable. All correlations above .02 are significant at the 95% confidence level. As expected, social acceptance was found to be positively correlated with GPA ($r = .13$, $p < .001$). Associations between school-level variables are presented in Table 3 along with estimated population means and standard deviations.

Multilevel Models

Longitudinal association between GPA and social acceptance. A series of models were estimated to address our hypotheses. Random effects for the intercept and slope were included in all models. Models 1 through 4 are presented in Table 4. Model 1 estimated the relation between grades and social support controlling for individual-level covariates. Based on previous research (Chen et al., 1997; Wentzel, 1991, 2005), we expected that, on average, there would be a positive association between GPA and social acceptance. In line with our prediction, GPA at Wave I did significantly predict positive changes in social acceptance.

Several demographic variables were also found to be predictive of changes in social acceptance: Men, higher SES adolescents, and younger adolescents all tended to have more positive changes in social acceptance over time. Additionally, Asians had less positive changes than Whites. The variance component associated with the random slope coefficient in Model 1 was also found to be significant ($p < .001$). This suggests that the relation between Wave I GPA and Wave II social acceptance does vary across schools, and provides justification for later models to look at school-level characteristics as moderators of this relation.

Ethnic group differences in the relation between GPA and social acceptance. Having established the direct association between GPA and social acceptance, Model 2 added the GPA by race interaction terms to test for ethnic group differences in social costs with academic success. Parameter estimates for Model 2 indicated that African American adolescents have a significantly weaker (more negative) relation between GPA and social acceptance than Whites. Specifically, while Whites (the reference group in Model 2) show a relatively strong positive association, the association for African Americans is negative, suggesting that there are differential social consequences with achievement for African Americans. While less Native Americans were present in the sample, as compared with Whites, they also showed significant social costs with achievement. Based on estimates from Model 2, Figure 1 shows the effect of Wave I GPA on subsequent changes in social acceptance for Whites, African Americans, and Native Americans. Although the relation between achievement and social acceptance was most negative for Native Americans, the difference between African Americans and Native Americans was not significant. Interaction terms for Hispanics and Asians in Model 2 were not significant, suggesting that the relation between grades and social acceptance for these groups tend, on average, to be the same as for Whites. In separate analyses, Hispanic and Asian subgroups were also

Table 4

Multilevel Parameter Estimates Showing the Effects of Grade Point Average (GPA) on Subsequent Changes in Social Acceptance as a Function of Ethnicity, Controlling for Individual- and School-Level Disadvantages

	B (*SE*)			
	Model 1	Model 2	Model 3	Model 4
Intercept	−.009 (.021)	−.013 (.022)	−.016 (.022)	−.010 (.024)
Level 1 predictors				
Social acceptance (Wave I)	.485 (.013)***	.484 (.013)***	.484 (.013)***	.483 (.014)***
Age	−.017 (.007)*	−.017 (.007)*	−.017 (.007)*	−.018 (.009)*
Male	.091 (.025)***	.090 (.024)***	.090 (.024)***	.090 (.024)
Black	.008 (.045)	.006 (.043)	.008 (.043)	.000 (.048)
Hispanic	−.089 (.046)	−.081 (.048)	−.079 (.048)	−.082 (.048)
Asian	−.136 (.058)*	−.124 (.068)	−.120 (.070)	−.125 (.070)
Native	−.044 (.106)	−.140 (.110)	−.138 (.110)	−.135 (.112)
Mixed	−.108 (.062)	−.099 (.061)	−.097 (.062)	−.099 (.063)
Other	−.247 (.145)	−.271 (.149)	−.260 (.148)	−.121 (.148)
SES	.032 (.012)*	.032 (.012)**	.031 (.012)*	.030 (.013)*
Single parent	−.061 (.033)	−.059 (.032)	−.061 (.032)*	−.057 (.032)*
GPA	.051 (.014)***	.070 (.016)***	.077 (.017)***	.092 (.020)***
Level 2 predictors				
SchSES				.015 (.019)
SchSafety				.007 (.023)
SchAchieve				−.001 (.018)
Log size				.001 (.021)
Not public				−.023 (.042)
Level 1 interactions				
GPA × Black		−.119 (.042)**	−.112 (.043)**	−.122 (.042)**
GPA × Hispanic		−.007 (.041)	.000 (.042)	−.018 (.045)
GPA × Asian		−.034 (.050)	−.040 (.049)	−.059 (.052)
GPA × Native		−.291 (.127)*	−.284 (.130)*	−.309 (.129)*
GPA × Mixed		.011 (.064)	.013 (.064)	−.001 (.063)
GPA × Other		.056 (.125)	.052 (.123)	.025 (.121)
GPA × SES			.008 (.013)	.008 (.014)
GPA × Single Parent			−.025 (.028)	−.030 (.030)
Cross-level interactions				
GPA × SchSES				.001 (.020)
GPA × SchSafety				−.015 (.021)
GPA × SchAchieve				.019 (.016)
GPA × Log Size				.005 (.011)
GPA × Not Public				.063 (.035)
Variance components				
Intercept	.0040 (.0018)*	.0039 (.0018)*	.0039 (.0017)*	.0033 (.0017)*
Slope (GPA)	.0064 (.0019)***	.0065 (.0020)***	.0064 (.0020)***	.0060 (.0019)***
Intercept–slope covariance	−.0031 (.0011)**	−.0031 (.0011)**	−.0029 (.0013)**	−.0031 (.0012)**
N (Level 1)	12,936	12,936	12,936	12,567
N (Level 2)	132	132	132	127

Notes. SES = socioeconomic status; SchSES = school-level SES; SchSafety = school-level safety; SchAchieve = school-level achievement.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

considered, but no differential social consequences were found.

Does disadvantage account for ethnic group differences in social costs? As African American and Native American adolescents are more likely than Whites to be from low-SES families, single-parent families, and disadvantaged schools, our next goal was to determine whether the significant interaction effect for African Americans and Native Americans might be accounted for by controlling for individual- and school-level disadvantage. Model 3 therefore added individual-level interactions

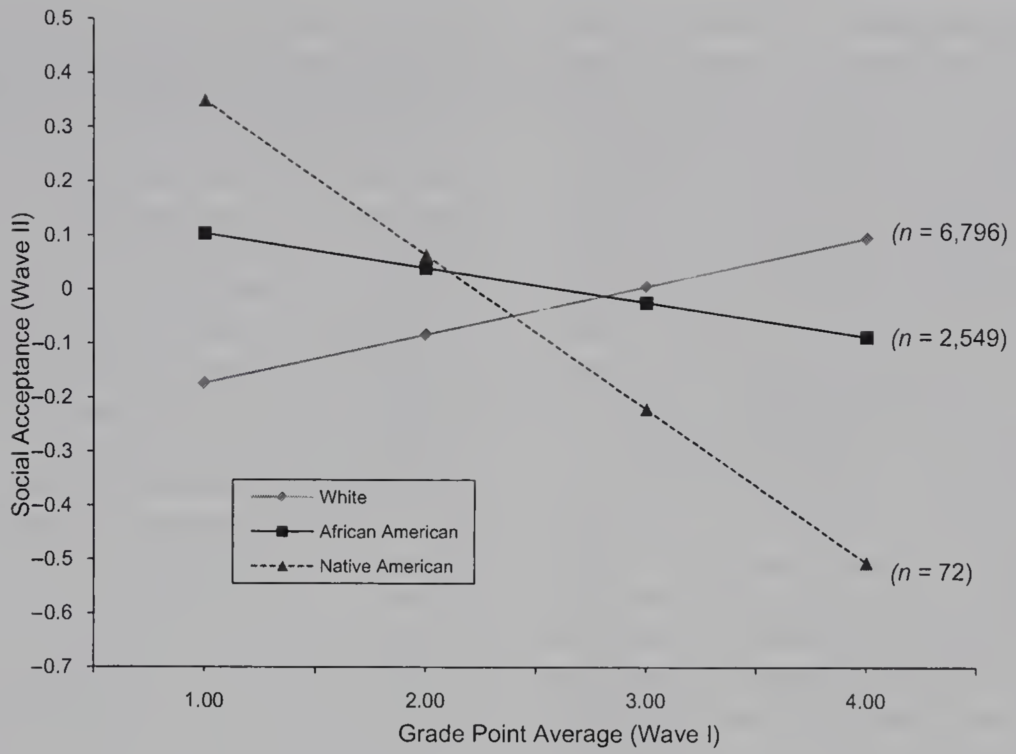


Figure 1. Fitted interaction plot depicting the relation between Wave I grade point average (GPA) and Wave II social acceptance for Black, Native American, and White adolescents.
Note. This figure is based on parameter estimates from Model 2 (Table 4), with GPA converted back to its original 4-point scale.

(GPA × SES; GPA × Single Parent) as competing moderators and found that although the GPA × Black and GPA × Native terms decreased slightly (by 6% and 2%, respectively), both effects remained significant. Model 4 then controlled for the effects of school disadvantage on the relation between grades and social acceptance. However, the GPA × Black interaction term remained highly significant, as did the GPA × Native term. Findings from these models suggest that individual- and school-level factors do not account for the differential social costs with academic success found for African Americans and Native Americans.

Is immigrant status a key variable influencing social costs? The next model considered whether less recent Hispanic, Asian, or Black immigrants show more social costs with academic success than their more recent immigrant counterparts. To test these effects, Model 5 included three-way interactions between GPA, race, and immigrant status. Parameter estimates for this model suggested that immigrant status did not play an important role for any of these groups. Specifically, with respect to the relation between grades and social acceptance, Hispanic adolescents with FBPs are not significantly different from Hispanic adolescents with U.S.-born parents (GPA × Hisp. × FBP: $B = -.080$, $SE = .115$, *ns*). Asian and Black adolescents also showed no differences in social costs with respect to immigrant status (GPA × Asian × FBP: $B = -.031$, $SE = .148$, *ns*; GPA × Black × FBP: $B = -.167$, $SE = .157$, *ns*). Unrelated to hypotheses for the current study, we also inadvertently discovered, in Model 5, that the main effect for Asians reported in Model 1, suggesting greater decreases in social acceptance, is being entirely driven by the effect for Asians with FBPs (Asian × FBP: $B = -.338$, $SE = .148$, $p < .05$). This suggests that Asian adolescents growing up in families with FBPs (first and second generations) tend to experience greater decreases in social acceptance across the adolescent period than Whites, while Asian adolescents from families with U.S.-born parents (third generation or greater) do not show greater decreases.

Does the social costs effect depend on gender? Model 6 tested for differences in social costs for African American men and women. Parameter estimates for this model suggest that no significant differences were present between men and women (GPA × Male × Black: $B = .040$, $SE = .065$, *ns*). Gender differences were also explored for each of the other ethnic groups and none were found to be significant. Models 5 and 6 (relating to immigrant status and gender) were not included as a table as a result of nonsignificant findings. A table detailing the parameter estimates for these models is available from the first author upon request.

School-level moderators of social costs for African Americans. The next set of models focused on exploring contingencies of the relation between grades and social acceptance for African Americans. (Because of the small sample size for Native Americans, it was not possible to consider differences across schools for this group.) Models 7 through 9 are presented in Table 5. Model 7 focused on the interaction between GPA, Black race, and school-level achievement to determine whether being in a higher achieving school is associated with more social costs for African Americans. Results for this model suggested that when considered alone, school-level achievement did not significantly affect the social costs of academic success for African Americans. Model 8 focused on the interaction between GPA, Black race, and Black school to determine whether being in a largely Black school might be protective against the social costs of academic success for African Americans. Results for this model also proved nonsignificant suggesting that, overall, the social costs of academic success for African Americans are not significantly different in largely Black schools.

Model 9 focused on testing whether the effects of being in a high-achieving school might depend on the proportion of Black students in the school or, in other words, whether it is important to consider school-level achievement and school racial context simultaneously. Parameter estimates from this model showed a significant interaction effect such that Black students in more high-achieving schools had greater social costs, but only when the school had a smaller percentage of Black students. This model therefore suggests that a more Black school context is protective against social costs when the school is high achieving.

With respect to model parameters, because the Black school variable (BlkSch) is dummy code, the significant GPA × Blk × SchAch term in Model 9 suggests that African American students in higher achieving non-Black schools have greater social costs with academic success than Black students in lower achieving non-Black schools ($B = -.232$, $SE = .098$, $p < .05$). The GPA × Blk × BlkSch × SchAch term shows that the increase in social costs as a result of being in a high-achieving school is almost completely eliminated if the school is largely Black ($B = .222$, $SE = .015$, $p < .05$). In other words, African Americans in higher achieving (+1 *SD* on school-level achievement), less Black schools are experiencing the greatest social costs with academic success, whereas African Americans in high-achieving Black schools are experiencing lower social costs. Figure 2 illustrates the school-level findings for four groups of Black students: Black students in *high-achieving non-Black schools*; Black students in *high-achieving Black schools*; Black students in *low-achieving non-Black schools*; and Black students in *low-achieving Black schools*. The slopes of the lines for Black students in low-achieving Black and non-Black schools are not significantly different from each other or from the slope of the line for Blacks in high-achieving Black schools. On the other hand, the slope of the line for Black students in high-achieving non-Black schools is significantly different from the others, indicating that Black students in these schools, on average, experience the greatest social costs with achievement.

School-level moderators of social costs for Hispanics. Additional models focused on the effects of context for Hispanics as well as for specific Hispanic ethnic groups. Specifically, our first set of models looked at the social consequences of achievement for Hispanic students as a function of school-level achievement and the percentage of Hispanic students in the school. Results of these analyses (Models 10, 11, and 12) suggested that school-level achievement and the proportion of Hispanic students in a school do not play a role for Hispanics.

We then considered whether school characteristics may have differential consequences for the Hispanic subgroups: Mexican, Puerto Rican, Cuban, and Central or South American. These analyses revealed that the school-context findings reported for Black students were replicated for Mexican students (Models 13, 14, and 15). Tables of the full models are not included owing to space limitations but are available from the first author upon request. Models 13 and 14 showed that school-level achievement and the percentage of Hispanic students in the school, when considered separately, did not significantly moderate the relation between achievement and social acceptance. However, when considered together, in Model 15, the same findings emerged as for African American students. Specifically, model parameters showed that Mexican students experience greater social costs with achievement in high-achieving schools ($B = -.236$, $SE = .107$, $p < .05$), but only when these schools do not have a substantial portion of Hispanic students ($B = .277$, $SE = .109$, $p < .05$). In other words, just as Black students experienced lower social costs with achievement in higher achieving Black schools, Mexican students also experienced lower social costs in higher achieving Hispanic schools. Figure 3 illustrates the school-level findings for students from Mexican descent. Specifically, Mexican students in higher achieving (+1 *SD* on school-level achievement) less Hispanic schools are shown to be experiencing the greatest social costs with

Table 5

Multilevel Parameter Estimates Showing the Effects of Grade Point Average (GPA) on Subsequent Changes in Social Acceptance as a Function of Race, School-Level Achievement, and the Proportion of Black Students in a School

	B (SE)		
	Model 7	Model 8	Model 9
Intercept	−.017 (.023)	−.009 (.023)	−.016 (.024)
Level 1 predictors			
Social acceptance (Wave I)	.483 (.014)***	.483 (.014)***	.483 (.014)***
Age	−.019 (.007)*	−.017 (.008)*	−.017 (.007)*
Male	.090 (.025)***	.092 (.025)***	.092 (.025)**
Black	.007 (.045)	.019 (.105)	.022 (.086)
Hispanic	−.083 (.048)*	−.081 (.050)	−.083 (.048)
Asian	−.124 (.067)	−.120 (.070)	−.124 (.068)
Native	−.134 (.111)	−.146 (.110)	−.138 (.112)
Mixed	−.097 (.061)	−.100 (.064)	−.010 (.062)
Other	−.256 (.149)	−.266 (.149)	−.258 (.149)
SES	.035 (.012)**	.034 (.012)**	.035 (.012)**
Single parent	−.053 (.033)	−.054 (.033)	−.051 (.033)
GPA	.074 (.017)**	.067 (.018)***	.072 (.019)***
Level 2 predictors			
BlkSch		−.016 (.038)	−.006 (.036)
SchAchieve	.017 (.012)		.017 (.013)
Level 1 interactions			
GPA × Black	−.117 (.044)**	−.182 (.108)	−.155 (.085)*
GPA × Hispanic	−.015 (.043)	−.011 (.042)	−.011 (.044)
GPA × Asian	−.037 (.048)	−.047 (.048)	−.040 (.045)
GPA × Native	−.305 (.123)*	−.293 (.128)*	−.302 (.123)*
GPA × Mixed	.004 (.064)	.005 (.064)	.002 (.064)
GPA × Other	−.048 (.124)	.054 (.126)	.048 (.124)
Level 2 and cross-level interactions			
GPA × BlkSch		.033 (.039)	.021 (.036)
GPA × SchAchieve	−.014 (.012)		−.016 (.012)
Black × BlkSch		−.021 (.119)	−.037 (.103)
Black × SchAchieve	−.043 (.041)		−.046 (.105)
BlkSch × SchAchieve			.006 (.036)
Black × BlkSch × SchAchieve			−.019 (.115)
GPA × BlkSch × SchAchieve			.023 (.040)
GPA × Black × BlkSch		.069 (.118)	.059 (.096)
GPA × Black × SchAchieve	−.042 (.032)		−.232 (.098)*
GPA × Black × BlkSch × SchAchieve			.222 (.105)*
Variance components			
Intercept	.0035 (.0018)*	.0037 (.0018)*	.0032 (.0017)*
Slope (GPA)	.0068 (.0020)***	.0063 (.0019)***	.0064 (.0019)***
Intercept–slope covariance	−.0034 (.0011)*	−.0030 (.0013)*	−.0031 (.0010)*
N (Level 1)	12,649	12,572	12,567
N (Level 2)	128	128	127

Note. SES = socioeconomic status; BlkSch = Black school (school in top quartile with respect to the percentage of Black students); SchAchieve = school-level achievement.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

achievement, whereas Mexican adolescents in other school contexts enjoy a positive association between achievement and social acceptance (as evidenced by the positive slopes for the other three lines). The slopes of the lines for Mexican students in low-achieving Hispanic and non-Hispanic schools are not significantly different from each other or from the slope of the line for Mexican students in high-achieving Hispanic schools. On the other hand, as indicated before, the slope of the line for Mexican

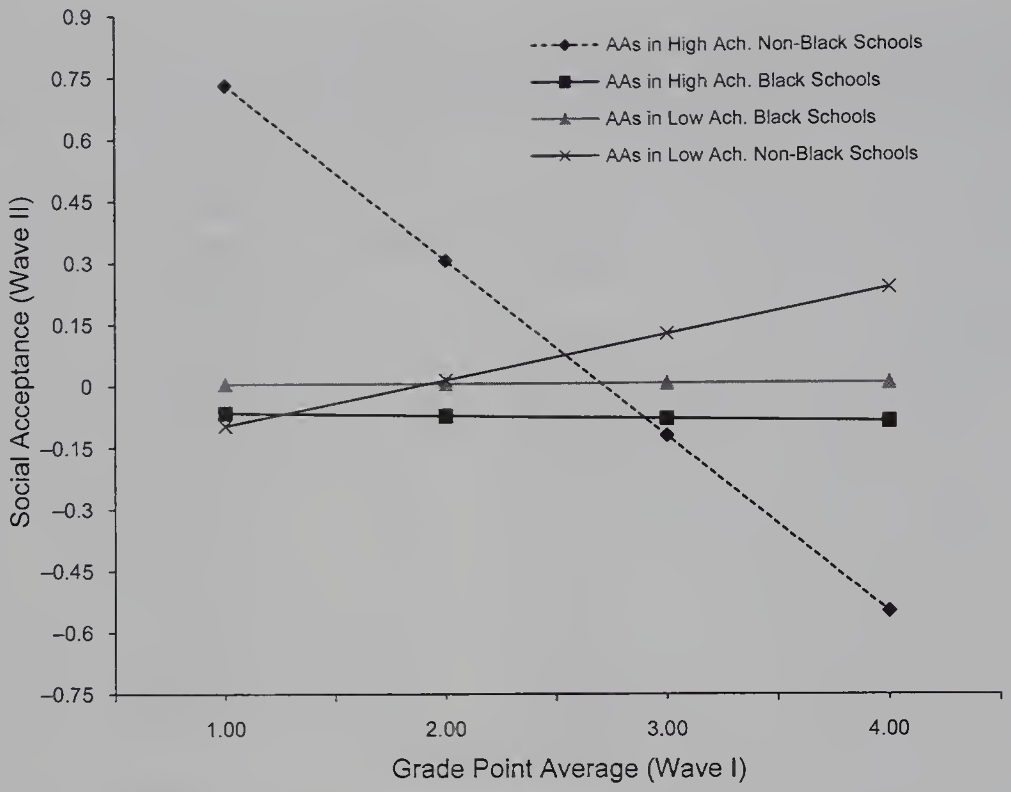


Figure 2. Fitted interaction plot depicting the relation between Wave I grade point average (GPA) and Wave II social acceptance for Black adolescents in four school contexts.
Note. This figure is based on parameter estimates from Model 9 (Table 5), with GPA converted back to its original 4-point scale. Sixty-eight percent of Black students (n = 1,708) are in Black schools (top quartile of percent Black), and 32% (n = 801) are in less Black schools. As school-level achievement is a continuous variable, specific sample sizes cannot be associated with each of the lines: high-achieving school = +1 *SD* from mean, low-achieving school = −1 *SD* from mean. AAs = African Americans.

students in high-achieving non-Hispanic schools is significantly different from the others, indicating that Mexican students in these schools, on average, experience the greatest social costs with achievement.

Other Hispanic ethnic groups were also considered across school contexts. Findings for Puerto Rican, Central or South American, and Cuban adolescents were, however, quite different. The percentage of Hispanic students was found not to influence the social costs of achievement for any of these groups. Furthermore, school-level achievement did not lead to increased social costs for any of these groups, regardless of whether or not the school was high achieving. In fact, for Puerto Rican and Central or South American adolescents school-level achievement was associated with lower social costs (more social acceptance) with achievement (GPA × Puerto Rican × SchAchieve: $B = .196$, $SE = .099$, $p < .05$; GPA × Central or South American × SchAchieve: $B = .124$, $SE = .045$, $p < .01$).

In additional analyses, the relation between achievement and social acceptance was considered for Black students as a function of the percentage of Hispanic students in the school and vice versa. No significant effects were found in these models, suggesting the importance of same-race minority peers, as opposed to minority peers in general. School context effects were also considered for Asians, but no significant differences were found. The percentage of White students in a school was also explored, but was not found to influence the effects of achievement on social acceptance for any of the groups considered.

Discussion

The current study explored the association between achievement and social acceptance across ethnic groups to test the proposition that some minority groups experience greater social costs with academic success. Furthermore, the effects of school-level variables on the relation between achievement and social acceptance were examined. To our knowledge, this study is the first to quantitatively demonstrate differential social costs with academic success across ethnic groups. In particular, results suggest that, as compared with Whites, Black and Native American adolescents experience more social costs with achievement, even after controlling for individual- and school-level disadvantage. These findings are in line with work

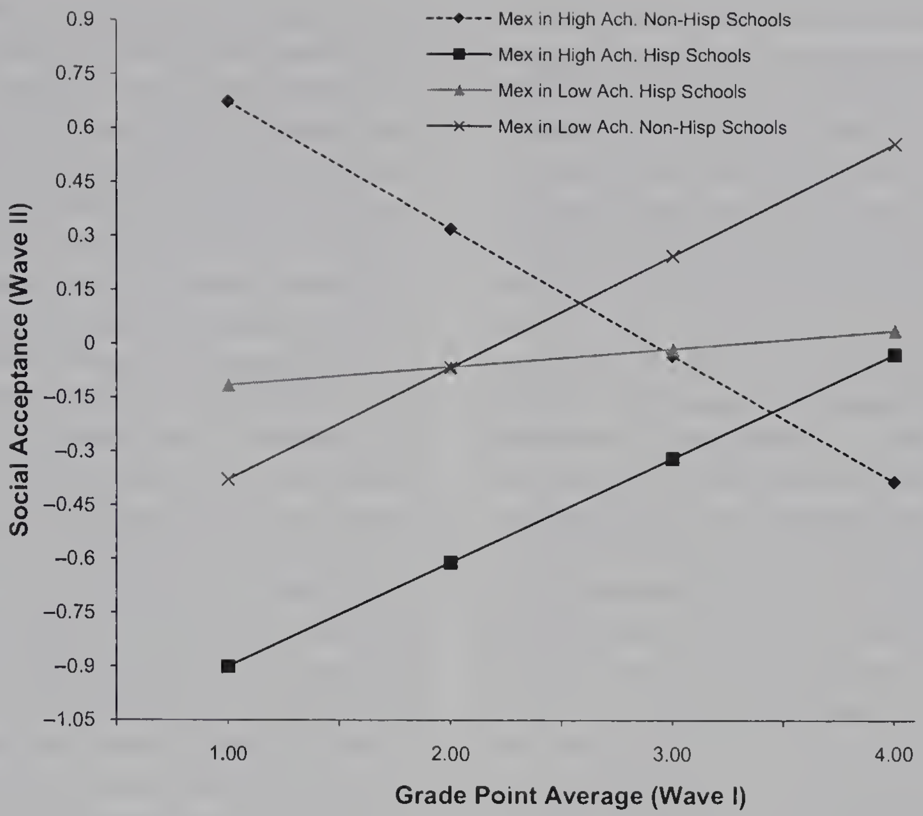


Figure 3. Fitted interaction plot depicting the relation between Wave I grade point average (GPA) and Wave II social acceptance for adolescents from Mexican descent in four school contexts.

Note. This figure is based on parameter estimates from Model 15, with GPA converted back to its original 4-point scale. Seventy-seven percent of Mexican students ($n = 689$) are in Hispanic schools (top quartile of percent Hispanic), and 23% ($n = 207$) are in less Hispanic schools. As school-level achievement is a continuous variable, specific sample sizes cannot be associated with each of the lines: high-achieving school = +1 *SD* from mean, low-achieving school = −1 *SD* from mean. Mex = adolescents from Mexican descent.

suggesting that greater social consequences may exist among currently stigmatized and historically oppressed groups (Fordham & Ogbu, 1986; Spencer & Harpalani, 2008; Steele, 1998).

An additional focus was to consider school-level contingencies. In particular, on the basis of existing work (e.g., Goldsmith, 2004), we explored school-level achievement and the proportion of same-race students in a school as moderators of the relation between achievement and social acceptance. Findings suggested that the social costs of achievement were greater for African Americans in higher achieving schools, but only when these schools had a smaller proportion of Black students. Specifically, for Black students, the social consequences of achievement were most severe in higher achieving schools with fewer Black students. However, when schools did have a substantial percentage of same-race peers, the social costs were relatively low, regardless of the levels of achievement within the school (Figure 1).

Consistent with our hypotheses, these findings suggest that the social consequences of achievement for African Americans are largely contingent on context (Cross, 2003; Spencer et al., 2001) and therefore are not likely to be the result of a widespread cultural orientation in opposition to achievement. Furthermore, results are in line with research suggesting that interracial tensions are more pronounced in competitive or high-achieving contexts (Brewer & Kramer, 1985), as well as with work suggesting that it is important for minority students to be exposed to a significant number of same-race peers (e.g., Goldsmith, 2004).

Our findings demonstrate that the greatest social costs with achievement for African American students are present in high-achieving schools with a smaller proportion of Black students. However, the results also show that, for Black students in predominantly Black schools, the relation between achievement and social acceptance is still not positive (see Figure 2). This is in contrast to a positive association between achievement and social acceptance for the majority group found in this study, as well as in previous work (Chen et al., 1997; Wentzel, 1991, 2005). The lack of positive association between achievement and social acceptance for African American adolescents in predominantly Black schools therefore suggests there are still some social costs with academic success for these students, relative to the majority group. This finding could be interpreted as in line with previous qualitative work conducted in predominantly Black schools (e.g. Fordham, 1988; Fordham & Ogbu, 1986). However, because the relation between achievement and social acceptance is more negative in competitive contexts, situational factors seem to be an important driving force behind this relation.

The school-level findings for African Americans were also replicated for students from Mexican descent. Specifically, Mexican students experienced substantially more social costs in higher achieving schools, but only when these schools did not contain a substantial portion of Hispanic students (Figure 3). However, other Hispanic groups did not show equivalent trends. In fact, the percentage of Hispanic students in a school had no effects for the Cuban, Puerto Rican, or Central or South American groups. Additionally, school-level achievement had no effect for Cubans and had a different effect for adolescents from Puerto Rican and Central or South American descent than for Mexicans or Blacks. Specifically, higher achieving schools were associated with lower social costs for Puerto Rican and Central or South American adolescents.

While this diversity in the social consequences of achievement among Hispanic groups was not predicted, a range of work is helpful in explaining these effects. First, relating to Mexican Americans, a variety of research suggests that with respect to subordination and disenfranchisement, the history of Mexican Americans is in many ways similar to that of Black Americans (Lopez & Stanton-Salazar, 2001; Portes & Rumbaut, 2001), and that the two groups are often similar in their reactions to discrimination (Portes, 1990). Given these perspectives, it is not surprising that Mexicans show similar tends to African Americans across school contexts. Several explanations are possible for the lack of social costs with achievement observed for Cuban, Puerto Rican, and Central or South American students across any of the contexts considered. Researchers have suggested that adolescents of Cuban descent tend to show fewer academic adjustment problems than those of Mexican descent (Portes, 1999). Additionally, most Cubans in our sample also tended to be in schools with a substantial portion of Hispanic students. With respect to Puerto Ricans, researchers have argued that their legal status as U.S. citizens may lead to less stigmatization and less problematic achievement dynamics than adolescents of Mexican descent (e.g., Flores-Gonzalez, 1999). Future work will be needed to test the various potential mechanisms behind the differential social costs experienced by particular groups. Overall, our findings are in line with the idea that acculturation experiences vary substantially across the individual ethnic groups within the Hispanic pan-ethnic category.

The current study also tested the possibility that immigrant status may be accounting for differences in social costs. Results of these analyses suggested that adolescents whose families have been in the United States for multiple generations did not show any greater social costs than their more recent immigrant counterparts. These findings, therefore, failed to provide support to the idea that length of time in the United States accounts for differences in social costs across groups (Portes & Zhou, 1993; Spencer et al., 2003). Based on previous research (Graham et al., 1998; Osborne, 1997), an additional hypothesis of the current study was that African American men might show greater social costs than women. Findings, however, showed no gender effect for African Americans, or for any of the other ethnic groups. This suggests that gender is not playing a direct role in the link between achievement and social acceptance.

Limitations and Future Directions

Although the findings of this study suggest that differential social costs do exist for particular groups, it was not within the scope of this research to determine the detailed mechanisms for these effects. It will therefore be important for future studies to develop a more detailed understanding of group differences, especially in school contexts where discrepancies in social costs across groups were found to be most pronounced. Work in this area has suggested that some differences in achievement values are present during early adolescence (e.g., Graham et al., 1998). However, the role of peer attitudes and norms in the relation between achievement and social acceptance is less understood. It will therefore be useful for future work to explore these effects, as well as to examine contextual factors that may be influencing the achievement values of minority peer groups (e.g., stereotypes and discrimination from the majority group, or socialization messages from families and communities).

Another area of future research will be to explore the role of identity in the social costs of achievement. This is of particular importance given that racial and ethnic identity are thought to influence perceptions of racism, as well as its consequences.

Furthermore, as differences in social costs across ethnic groups may be the result of attitudes and norms within peer groups, network techniques, which aggregate the characteristics of peers (e.g., Kiesner, Cadinu, Poulin, & Bucci, 2002), will be necessary to test these effects. Contextual factors such as socialization and experiences with discrimination will also be important to consider—particularly in relation to their role as predictors of group identity (Hughes et al., 2006; Sellers & Shelton, 2003).

Finally, it is unclear from the current study whether the social consequences of achievement experienced by particular groups come from same-race peers, from members of the dominant group, or from elsewhere. Distinguishing the source of social consequences with achievement will therefore be an important issue for future studies to address. It will be helpful to consider the possibility that greater social costs experienced by particular groups are the result of increased hostility from the majority group at higher levels of achievement. Specifically, given that research suggests that Whites direct their prejudice unequally, as a function of individual characteristics (Kaiser & Pratt-Hyatt, 2009), it is plausible that White adolescents may feel particularly threatened when members of specific groups achieve academically and therefore may express more hostility toward high-achieving members of these groups. Future research will be necessary to distinguish these various possible sources of social costs.

Conclusion

In closing, we wish to emphasize that although our findings indicate that ethnic group differences do exist, it is by no means our intention to insinuate that such differences are set in stone. Quite the opposite: Our hope is that a more nuanced understanding of ethnic group differences, and the complex dynamics behind them, will stimulate actions to address these issues within education systems at multiple levels. Although many researchers have argued for a color-blind perspective, suggesting that ethnic issues are irrelevant when background and contextual factors are considered (e.g., Cook & Ludwig, 1997; Rothstein, 2004), results of the current study suggest strongly otherwise. An important implication of this article is therefore that to redress ethnic gaps in achievement, in addition to focusing on socioeconomic and structural problems, schools and communities will need to understand and address issues of race.

Furthermore, while our results suggest that schools with more same-ethnic peers may be protective, we do not wish to imply that segregated schools are a necessity for minority students to maintain healthy social relations alongside academic achievement. We feel that a more appropriate implication of these findings is that schools with characteristics associated with differential social costs should seek to develop a greater awareness of unacknowledged stereotypes and take additional measures to maintain healthy racial dynamics. Specifically, previous recommendations for school personnel to focus on earning and maintaining the trust of minority students and to create an environment of "identity safety" (Davies et al., 2005; Ogbu & Simons, 1998) may be of particular importance in high-achieving schools where stigmatized minority groups are a clear minority.

References

Ainsworth-Darnell, J. M., & Downey, D. B. (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review*, *63*, 536–553.

Akom, A. A. (2003). Reexamining resistance as oppositional behavior: The Nation of Islam and the creation of a Black achievement ideology. *Sociology of Education*, *76*, 305–325.

Baumeister, R., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529.

Brewer, M., & Kramer, R. (1985). The psychology of intergroup attitudes and behaviors. *Annual Review of Psychology*, *36*, 219–243.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, *99*, 353–395.

Chantala, K., & Tabor, J. (1999). *Strategies to perform a design-based analysis using the Add Health data*. Chapel Hill: Carolina Population Center, University of North Carolina.

Chen, X., Rubin, K. H., & Li, D. (1997). Relation between academic achievement and social adjustment: Evidence from Chinese children. *Developmental Psychology*, *33*, 518–525.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.

Cook, P. J., & Ludwig, J. (1997). Weighing the burden of "Acting White": Are there race differences in attitudes towards education? *Journal of Policy Analysis and Management*, *16*, 256–278.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Crosnoe, R., Cavanagh, S., & Elder, G. H. (2003). Adolescent friendships as academic resources: The intersection of friendship, race, and school disadvantage. *Sociological Perspectives*, *46*, 331–352.

Cross, W. E. (1991). *Shades of black*. Philadelphia, PA: Temple University Press.

Cross, W. E. (2003). Tracing the historical origins of youth delinquency and violence: Myths and realities about Black culture. *Journal of Social Issues*, *59*, 67–82.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, *88*, 276–287.

Etzkowitz, H., Kernelgor, C., Neuschatz, M., Uzzi, B., & Alonzo, J. (1994). The paradox of critical mass for women in science. *Science*, *266*, 51–54.

Finch, B. K., Kolody, B., & Vega, W. (2000). Perceived discrimination and depression among Mexican-origin adults in California. *Journal of Health and Social Behavior*, *4*, 295–313.

Flores-Gonzalez, N. (1999). Puerto Rican high achievers: An example of ethnic and academic identity compatibility. *Anthropology and Education Quarterly*, *10*, 343–362.

Fordham, S. (1988). Racelessness as a factor in Black students' school success: A pragmatic strategy or pyrrhic victory? *Harvard Educational Review*, *58*, 54–84.

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the "burden of 'acting White'." *The Urban Review*, *18*, 176–206.

Goldsmith, P. A. (2004). Schools' racial mix, students' optimism, and the Black-White and Latino-White achievement gaps. *Sociology of Education*, *77*, 121–148.

Gonzales, N. A., Cauce, A. M., Friedman, R. J., & Mason, C. A. (1996). Family, peer; and neighborhood influences on academic achievement among African-American adolescents. *American Journal of Community Psychology*, *24*, 365–387.

Graham, S., Taylor, A., & Hudley, C. (1998). Exploring achievement values among ethnic minority early adolescents. *Journal of Educational Psychology*, *90*, 606–620.

Hagerty, B. M. K., & Patusky, K. L. (1995). Developing a measure of sense of belonging. *Nursing Research*, *44*, 9–13.

Hanushek, E. A., & Taylor, L. L. (1990). Alternative assessments of the performance of schools: Measurement of state variations in achievement. *Journal of Human Resources*, *25*, 179–201.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford.

Hughes, D., Rodriguez, J., Smith, E. P., Johnson, D. J., Stevenson, H. C., & Spicer, P. (2006). Parents' ethnicracial socialization practices: A review of research and directions for future study. *Developmental Psychology*, *42*, 747–770.

Johnson, M. K., Crosnoe, R., & Elder, G. H. (2001). Students' attachment and academic engagement: Role of race and ethnicity. *Sociology of Education*, *74*, 318–340.

Kaiser, C. R., & Pratt-Hyatt, J. S. (2009). Distributing prejudice unequally: Do Whites direct their prejudice toward strongly identified minorities? *Journal of Personality and Social Psychology*, *96*, 432–445.

Kiesner, J., Cadinu, M., Poulin, F., & Bucci, M. (2002). Group identification in early adolescence: Its relation with peer adjustment and its moderator effect on peer influence. *Child Development*, *73*, 200–212.

Lopez, D., & Stanton-Salazar, R. (2001). Mexican Americans: A second generation at risk. In R. G. Rumbaut & A. Portes (Eds.), *Ethnicities: Children of immigrants in America* (pp. 57–90). Los Angeles: University of California Press.

McLoyd, V. C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioemotional development. *Child Development*, *61*, 311–346.

O'Connor, C., Fernandez, S. D., & Girard, B. (2007). The meaning of "Blackness": How Black students differentially align race and achievement across time and space. In A. Fuligni (Ed.), *Contesting stereotypes and creating identities: Social categories, identities and educational participation* (pp. 91–114). New York: Russell Sage Foundation.

Ogbu, J. (1999). Beyond language: Ebonics, proper English, and identity in a Black-American speech community. *American Educational Research Journal*, *36*, 147–184.

Ogbu, J. (2004). Collective identity and the 'burden of acting White' in black history, community, and education. *Urban Review*, *36*, 1–35.

Ogbu, J., & Simons, H. D. (1998). Voluntary and involuntary minorities: A cultural-ecological theory of school performance with some implications for education. *Anthropology and Education Quarterly*, *29*, 155–188.

Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, *89*, 728–735.

Paige, R. (2004). *Paige calls achievement gap "major driver of racial inequity" in the United States*. U.S. Department of Education Press Release. Retrieved April 6, 2008, from the World Wide Web: http://www.ed.gov/news/pressreleases/2004/07/07222004.html

Perie, M., & Moran, R. (2005). *NAEP 2004 Trends in academic progress: Three decades of student performance in reading and mathematics (NCES 2005-464)* (U.S. Department of Education). Washington, DC: Government Printing Office.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, *108*, 499–514.

Pong, S. (1998). The school compositional effect of single parenthood on 10th-grade achievement. *Sociology of Education*, *71*, 23–42.

Portes, A. (1990). From south of the border: Hispanic minorities in the United States. In V. Yans-McLaughlin (Ed.), *Immigration reconsidered: History, sociology, and politics* (pp. 160–186). New York: Oxford University Press.

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. New York: Russell Sage Foundation.

Portes, A., & Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *Annals of the American Academy of Political and Social Science, 530*, 74–96.

Portes, P. R. (1999). Social and psychological factors in the academic achievement of children of immigrants: A cultural history puzzle. *American Educational Research Journal, 36*, 489–507.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual* (Tech. Rep. No. 2001/01) [Software manual]. London: Department of Biostatistics and Computing, University of London.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the Black-White achievement gap*. Washington, DC: Economic Policy Institute.

Russell, D. W. (1996). The UCLA Loneliness Scale (Version 3): Reliability, validity and factorial structure. *Journal of Personality Assessment, 66*, 20–40.

Sellers, R. M., & Shelton, J. N. (2003). The role of racial identity in perceived racial discrimination. *Journal of Personality and Social Psychology, 84*, 1079–1092.

Spencer, M. B. (1999). Social and cultural influences on school adjustment: The application of an identity-focused cultural ecological perspective. *Educational Psychologist, 34*, 43–57.

Spencer, M. B., Cross, W. E., Jr., Harpalani, V., & Goss, T. N. (2003). Historical and developmental perspectives on Black academic achievement: Debunking the ''acting White'' myth and posing new directions for research. In C. C. Yeaky & R. D. Henderson (Eds.), *Surmounting all odds: Education, opportunity, and society in the new millennium* (pp. 273–304). Greenwich, CT: Information Age Publishers.

Spencer, M. B., & Harpalani, V. (2008). What does ''acting White'' actually mean? Racial identity, adolescent development, and academic achievement among African American youth. In J. U. Ogbu (Ed.), *Minority status, oppositional culture and schooling* (pp. 543–601). Mahwah, NJ: Erlbaum.

Spencer, M. B., & Markstrom-Adams, C. (1990). Identity processes among racial and ethnic minority children in America. *Child Development, 61*, 290–310.

Spencer, M. B., Noll, E., Stoltzfus, J., & Harpalani, V. (2001). Identity and school adjustment: Questioning the ''Acting White'' assumption. *Educational Psychologist, 36*, 21–30.

StataCorp. (2007). Stata statistical software (Release 10) [Software]. College Station, TX: StataCorp.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.

Steele, C. M. (1998). Stereotyping and its threats are real. *American Psychologist, 53*, 680–681.

Tyson, K. (2002). Weighing in: Elementary-age students and the debate on attitudes toward school among Black students. *Social Forces, 80*, 1157–1189.

Tyson, K., Darity, W., & Castellino, D. (2005). ''It's not 'a Black thing''': Understanding the burden of acting White and other dilemmas of high achievement. *American Sociological Review, 70*, 582–605.

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development, 62*, 1066–1078.

Wentzel, K. R. (2005). Peer relationships, motivation, and academic performance at school. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 279–296). New York: Guilford.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*, 1–37.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1714–1728

Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Identity

Thomas D. Lyon and
Elizabeth C. Ahern
University of Southern California

Lindsay C. Malloy
University of Cambridge

Jodi A. Quas
University of California, Irvine

A total of two hundred ninety-nine 4- to 9-year-old maltreated and nonmaltreated children of comparable socioeconomic status and ethnicity judged whether children should or would disclose unspecified transgressions of adults (instigators) to other adults (recipients) in scenarios varying the identity of the instigator (stranger or parent), the identity of the recipient (parent, police, or teacher), and the severity of the transgression (''something really bad'' or ''something just a little bad''). Children endorsed more disclosure against stranger than parent instigators and less disclosure to teacher than parent and police recipients. The youngest maltreated children endorsed less disclosure than nonmaltreated children, but the opposite was true among the oldest children. Older maltreated children distinguished less than nonmaltreated children between parents and other types of instigators and recipients.

What happens when children witness adult transgressions? Do children believe that transgressions should be disclosed or do they believe that the acts should be kept secret? Are children's attitudes affected by adverse familial experiences? Do children's beliefs vary across age, different transgressor identities, types of transgressions, and the person to whom the transgression is reported? Although these questions are of considerable theoretical and practical interest, there is virtually no research examining the development of attitudes regarding disclosure of adult transgressions. This study examined those attitudes among 4- to 9-year-old maltreated children and nonmaltreated children of comparable socioeconomic status and ethnicity.

Several lines of research are relevant to understanding children's emerging attitudes about disclosing adult transgressions: (a) children's disclosure of maltreatment and the effects of maltreatment on children's attitudes; (b) children's attitudes about disclosing the transgressions of others, typically peers; (c) children's attitudes about lying and children's honest and dishonest behavior; and d) children's attitudes about authority and the validity of authoritative demands. Each is discussed in turn. To facilitate comparisons across studies, we refer to persons who commit misdeeds and may encourage transgression secrecy as instigators, and those to whom the child might disclose transgressions as recipients.

Disclosure of Childhood Maltreatment

A real-world example of childhood disclosure of adult transgressions concerns the disclosure of child abuse, and a small but important body of

---

Portions of this research were presented at the American Psychology-Law Society, Jacksonville, FL, 2008, and the third annual meeting of the Conference on Empirical Legal Studies, Ithaca, NY, 2008. Vera Chelyapov, Megan Donahue, Hilary Feybush, Laura Hinojosa, Ektha Aggarwal, Deanna Cano, Jennie King, Kathryne Krueger, Irma Escobar, Bruce John, Emily Kerr, Crystal Lara, Leah Manchester, Cindy Mekari, Deborah Ogbonna, Shadie Parivar, Megan Sim, and LeighAnn Smith assisted in data collection and coding. We are grateful for the support of the Presiding Judge of the Los Angeles County Juvenile Court, the Los Angeles County Department of Children's and Family Services, Los Angeles County Children's Law Center, the Children's Services Division of Los Angeles County Counsel, and the Los Angeles County Child Advocate's Office. Preparation of this article was supported in part by National Institute of Child Health and Human Development Grant HD047290-01A2 and a Doctoral Dissertation Fellowship from the National Science Foundation (SES-0720421).

Correspondence concerning this article should be addressed to Thomas D. Lyon, University of Southern California Law School, 699 Exposition Blvd., Los Angeles, CA 90089-0071. Electronic mail may be sent to tlyon@law.usc.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0008

research has examined the process of disclosure among child victims of abuse. Findings offer some insight into how the identity of the instigator and recipient may influence children's attitudes about disclosure and highlight the need to examine attitudes in both maltreated and nonmaltreated samples. Maltreated children are often instructed to keep abuse a secret (Bussey & Grimbeek, 1995; Lyon, 2002), and surveys of adults reporting that they were sexually abused as children have found that a majority report never having disclosed the abuse during childhood (Laumann, Gagnon, Michael, & Michaels, 1994).

For one, the relationship of the child to the instigator affects the likelihood that the child will disclose: Children are less likely to disclose abuse by parents than by strangers (London, Bruck, Wright, & Ceci, 2008; Lyon, 2010). Furthermore, parents (typically mothers) are usually the first recipients when children disclose abuse (Anderson, Martin, Mullen, Romans, & Herbison, 1993; Fleming, 1997), especially among children younger than 10 years of age (Hershkowitz, Lanes, & Lamb, 2007; Kogan, 2004). Parental support has been shown to have an effect on the likelihood that children report abuse (Lawson & Chaffin, 1992), delay reporting (Hershkowitz et al., 2007), or maintain their report over time (Malloy, Lyon, & Quas, 2007).

Abuse by parents is less likely than abuse by strangers to be reported to the police (Hanson, Resnick, Saunders, Kilpatrick, & Best, 1999). Indeed, disclosure of sexual abuse to any official authorities (e.g., police or social workers) is quite rare; respondents in adult surveys state that only about 10% of abuse was reported to authorities (Mullen, Martin, Anderson, Romans, & Herbison, 1993; Russell, 1983; Smith et al., 2000).

If these patterns reflect children's reasoning about disclosure, the findings suggest that children will endorse more disclosure of wrongdoing to parents than to the police or teachers, particularly when the instigator is another parent. However, the low rates of disclosure to authorities may reflect caretakers' attitudes about disclosure rather than children's, since caretakers are most likely to control the decision regarding whether to report the abuse. Indeed, the extent to which children disclose to parents rather than others may simply reflect children's dependency on their parents rather than a clear preference for their parents as recipients.

The attitudes of maltreated children are of special interest, given the importance of maltreatment disclosure and the potential that maltreatment itself affects children's attitudes about transgression secrecy. Maltreated children have been found to make similar moral judgments as nonmaltreated children (Smetana, Daddis, et al., 1999; Smetana, Kelly, & Twentyman, 1984; Smetana, Toth, et al., 1999). For example, when asked about transgressions, both maltreated and nonmaltreated young children tended to classify moral violations (physical harm, psychological harm, or inequities in resources) as "very bad" and social conventional violations (disobeying school rules) as "a little bad" (Smetana et al., 1984). However, maltreated children are more likely to expect that others, including caretakers, will behave in unsupportive, rejecting, or punitive ways (Macfie et al., 1999; Shields, Ryan, & Cicchetti, 2001; Shipman & Zeman, 2001; Shipman, Zeman, Fitzgerald, & Swisher, 2003; Toth, Cicchetti, Macfie, Rogosch, & Maughan, 2000).

If children's expectations of the reactions of others affect their reasoning about disclosure, then maltreated children may reason differently about the disclosure of adult transgressions than nonmaltreated children. On the one hand, fears of the instigator and distrust of the recipient may make them less inclined to endorse disclosure in general. On the other hand, their negative attitudes about parents may make them less protective of parents and thus less likely to distinguish between stranger and parent instigators when endorsing disclosure.

Children's Age-Related Attitudes About Disclosing Transgressions

Although a few studies have examined children's developing attitudes about reporting wrongdoings, findings are inconsistent. Some research suggests that children are less likely to endorse disclosure of wrongdoing with age: Whereas most 5- to 6-year-olds endorsed disclosure, 8- to 10-year-olds were more inclined to endorse secrecy (Piaget, 1932; Watson & Valtin, 1997). Other research, however, suggests the opposite: increased disclosure endorsement with age. An unpublished study by Wilson and Pipe (reported in Pipe & Goodman, 1991) found that 9-year-olds were more likely to endorse disclosure of accidental breakage of a toy than 6-year-olds, and Rotenberg (1995) found some evidence that 9-year-olds were more likely than younger children to endorse disclosure of their own transgressions.

One reason for the variable results concerns the identity of the instigator. In studies finding decreased disclosure endorsement with age, the instigator was always a peer (either a brother, Piaget, 1932; or friend, Watson & Valtin, 1997). The age-related decrease could be owing to increasing protectiveness of peers, akin to children's greater dependency on peer relationships as they approach adolescence (Buhrmester & Prager, 1995; Hunter & Youniss, 1982). In contrast, research that found increased endorsement of disclosure in this age range either left the instigator unspecified (Pipe & Goodman, 1991) or identified the child herself as the instigator (Rotenberg, 1995). Children might judge disclosure of adult transgressions quite differently. Similarly, the research has generally failed to vary the identity of the recipient. With the exception of Watson and Valtin (1997), who found that children were more likely to endorse disclosure to a mother than a friend, the identity of the recipient was either unspecified (e.g., ''adults''; Rotenberg, 1995; Pipe & Goodman, 1991) or a parent (Piaget, 1932). Because children's attitudes may differ depending on the identity of the instigator and recipient, it is essential to evaluate the effects of such identities directly as well as in conjunction with children's age and maltreatment status.

Children's Attitudes About Obedience to Authority

Developmental changes in children's beliefs about obedience to authority may further influence their beliefs about disclosing transgressions. Young children, for example, may view disclosure of adult transgressions as an act of disobedience, particularly if the transgressor instructs them not to disclose. Piaget (1932) found evidence that before about 8 years of age, children are ''moral realists'' and believe that ''any act that shows obedience to a rule or even to an adult, regardless of what he may command, is good'' (p. 111). Subsequent research has questioned Piaget's conclusions, finding that children at all ages judge that authoritative adults cannot justly order children to engage in conduct that is clearly harmful to the child or others (Damon, 1977; Laupa & Turiel, 1986). The difference in results may partly be attributable to the types of conduct studied; whereas Piaget's scenarios concerned the principle of equality of duties (children sharing equally in chores), subsequent research has examined more serious transgressions, such as stealing (Damon, 1977) or physical fighting (Laupa & Turiel, 1986).

If children do not view secrecy about wrongdoing as unambiguously immoral, their attitudes may be more subject to authoritative commands. Indeed, when children as young as 4 have been questioned, nontrivial percentages will state that parents and teachers can order children to behave immorally, both with respect to lying (Laupa, Turiel, & Cowan, 1995) and fighting (Laupa, 1994). Hence, if young children in particular view nondisclosure as deferential to authority, then one would expect these children to be more reluctant to endorse disclosure, whereas with age, children's disclosure endorsement should increase.

Children's Attitudes About Lying

In contrast to the scant research on children's attitudes about transgression secrecy, a substantial amount of research has examined children's attitudes about lying. By at least 4 years of age, children believe that lying is wrong (Bussey, 1992; Peterson, Peterson, & Seeto, 1983) and are more critical of lies that conceal transgressions than white lies (lies not intended to do harm; Bussey, 1999; Peterson et al., 1993), and these attitudes do not change with age. Maltreated children develop similar attitudes (Lyon & Dorado, 2008; Lyon & Saywitz, 1999), and there is some evidence that young maltreated children are more sensitive than nonmaltreated children to the wrongfulness of lying (Lyon, Carrick, & Quas, 2010).

However, most research to date has concerned children's attitudes about concealing self-transgressions or other children's wrongdoing. In one of the few studies that investigated children's attitudes about lying for adults, Wagland and Bussey (2005) asked 5- to 10-year-olds to predict whether a child witnessing an adult friend commit a transgression would lie to his or her father. Although children's predictions of honesty were influenced by whether the father reassured the child that punishment was not forthcoming or emphasized the advantages of telling the truth, children never showed a tendency to predict lying, even in scenarios in which the child was implicated in the wrongdoing, the adult friend warned against disclosure, and the father emphasized the fact that the child would be punished.

Wagland and Bussey (2005) did not vary the identity of the instigator or the recipient. The instigator was always an adult friend and the recipient was always the child's father. Yet, there are several reasons to suspect that instigator and recipient identities would play a role in influencing children's attitudes about lying. For one, children are more likely to lie for a parent than for a stranger (Tye, Amato, Honts, Devitt, & Peters, 1999). Instigator warnings increase children's willingness to lie (Bussey, 1993). Second, although recipient identity has not been manipulated, recipient actions can increase honesty, such as by eliciting a promise to tell the truth (Lyon & Dorado, 2008; Lyon, Malloy, Quas, & Talwar, 2008; Talwar, Lee, Bala, & Lindsay, 2002, 2004), reassuring the child that disclosure will not get the child into trouble (Lyon & Dorado, 2008; Lyon et al., 2008), and discussing the importance of truthfulness (Huffman, Warren, & Larson, 1999; London & Nunez, 2002). Recipient identity may play a role in children's endorsement of transgression disclosure to the extent that children expect different reactions from different types of recipients.

However, it is unclear whether findings concerning children's honest and deceptive behavior are directly applicable to children's attitudes about transgression disclosure. Children may reason about secrecy differently than they reason about lying. Secrets, unlike lies, constitute omission of the truth rather than commission of a falsehood. Ethicists have noted that although one might maintain that lies are presumptively immoral, one cannot make the same claims about secrecy (Bok, 1982). Furthermore, research on children's lying behavior may not reflect children's reasoning: Children's behavior is likely the product of both deliberation about the morality of lying and the weighing of other factors (such as the likelihood of detection and punishment; Pipe & Wilson, 1994). Nevertheless, the research on children's attitudes about lying suggests that children will be more inclined to endorse disclosure of wrongdoing when the instigator is a stranger, and when the recipient is a person they trust, such as a parent.

Present Study

The purpose of this study was to examine maltreated and nonmaltreated children's attitudes about adult transgression secrecy. We presented 4- to 9-year-old children with and without a history of substantiated child maltreatment with scenarios involving one adult "instigator" (an adult wrongdoer who demands secrecy) and one adult "recipient" (an adult to whom the disclosure might be made). We varied the identity of the instigator (parent or stranger) and the identity of the recipient (parent, police, or teacher). We also varied the severity of the transgression, describing the adult's actions as "just a little bad" or "really bad." We left the acts unspecified to avoid variability in children's responses to specific factual scenarios, although, consistent with prior work (Smetana et al., 1984), we expected children to understand that the actions, particularly those that were "really bad," constituted moral violations. Finally, we varied whether children were asked if the story child "should" or "would" disclose the wrongdoing, to uncover potential differences between judgments about the morality of disclosure (as revealed by what "should" occur) and attitudes about the actual likelihood of disclosure (as revealed by what "would" occur).

Several hypotheses were advanced. First, with respect to maltreatment, we hypothesized that maltreated children would endorse less disclosure overall than nonmaltreated children because of their expectation that adults are unsupportive. We also predicted maltreated children would discriminate less among different types of instigators and recipients because they would be less inclined to view parents as deserving of special consideration or trust. Second, given the inconsistent age differences in former studies, we tentatively predicted that older children would endorse more disclosure of adult transgressions than younger children because of children's diminishing deference to authority.

Third, with respect to the child's relationship to the instigator and recipient, we predicted that children would be more likely to endorse disclosure against strangers than parents, and more likely to endorse disclosure to parents than police and teachers. We also expected to see some evidence of keeping disclosure within the family (so that there would be an interaction between instigator and recipient: Parent–parent disclosure would be endorsed more often than parent–nonparent disclosure). Fourth, we expected children to endorse more disclosure when asked what the story child "should" do than when asked what the story child "would" do because the former would elicit children's moral judgments whereas the latter might elicit attitudes about behavior influenced by more than morality (such as consequences). Fifth, we expected children to endorse more disclosure when the transgression was described as "really bad" rather than "just a little bad" analogous to children's differentiation of white lies versus harmful lies (Bussey, 1999; Peterson et al., 1993), and of violations of social convention versus moral violations (Smetana et al., 1984).

Finally, we advanced several predictions concerning children's explanations for their responses concerning transgression disclosure. Specifically, we expected that, in general, children's explanations would track their patterns of disclosure endorsement. Accordingly, for children who endorsed disclosure, we predicted that there would be more references to the instigator in the stranger than parent instigator stories (because the identity of the instigator would justify disclosure in the stranger stories) and the opposite pattern for children who endorsed nondisclosure. With respect to transgression severity, we expected that children who endorsed disclosure would mention the transgression more often when the transgression was ''really bad'' than when it was ''just a little bad,'' and the opposite pattern for children who endorsed nondisclosure. We further expected that children would refer more often to the instigator's command (don't tell) when endorsing nondisclosure than when endorsing disclosure.

Method

Participants

The sample consisted of two hundred ninety-nine 4- to 9-year-old maltreated ($N = 153$) and nonmaltreated children ($N = 146$; $M = 6$ years 7 months, $SD = 1.71$; 54% girls). The sample was ethnically diverse (62% Latino, 27% African American, 6% Caucasian, 5% Other). The maltreated sample consisted of children substantiated as suffering from child neglect and/or physical or sexual abuse who had been removed from the custody of their parents or guardians. Children gave their assent to participate, and consent was obtained from the Presiding Judge of Juvenile Court and the Children's Law Center, which represents children in dependency proceedings. Children were not eligible if they were awaiting an adjudication hearing or a contested disposition hearing on the date of testing (because they might be asked to testify), if interpreter services were provided to their family and they were clearly incapable of communicating with the researchers in English, or if their attorney had objected to their participation. The nonmaltreated sample consisted of children recruited from schools serving predominantly low-income ethnic minority families in neighborhoods comparable to those from which most maltreated children were removed. Children in the nonmaltreated sample who were not in the custody of one or both parents were excluded because of the potential that they had been removed from their parent's care because of maltreatment.

Materials and Procedures

Rapport building. The experimenter first built rapport with the child using questions adapted from those described in the National Institute of Child Health and Human Development Investigative Interview Protocol (Lamb, Hershkowitz, Orbach, & Esplin, 2008; Orbach et al., 2000). The child was asked about things he or she liked and did not like to do.

Disclosure vignettes. The experimenter read 20 brief vignettes (accompanied by pictures of characters) in which an adult (instigator) committed a misdeed and told a story child not to disclose. The story child was then depicted with a different adult (recipient) and the experimenter asked the child if the story child ''should'' or ''would'' tell the recipient what the instigator did, and then asked for an explanation. A sample vignette is:

> Here is a boy. Here is his dad. His dad did something really bad. The dad says to the boy, ''Don't tell.'' Now the boy is with his mom. Should the boy tell his mom what his dad did? Why?

The gender of the story child always matched the gender of the subject. Within subjects, the scenarios varied with respect to the instigator's identity (father, mother, or stranger, the latter defined as ''someone [the story child] does not know''), the recipient's identity (father, mother, police officer, or teacher), and the severity of the wrongdoing (''really bad'' or ''just a little bad''). These were fully crossed except that there were no vignettes in which the instigator and the recipient were the same (e.g., no father instigator–father recipient vignettes). The order of scenario administration was counterbalanced. Half of the children heard vignettes asking what the story child ''should'' do and half heard vignettes asking what the story child ''would'' do.

Verbal ability. At the end of the procedure children were administered the Productive Vocabulary subsection of the Woodcock–Johnson Psycho-Educational Battery–Revised: Tests of Cognitive Ability (WJ–R; Woodcock & Johnson, 1989) to assess verbal ability.

Results

Comparability of Maltreated and Nonmaltreated Children

The maltreated and nonmaltreated children were comparable in age (maltreated: $M = 6$ years 8 months, $SD = 1.60$; nonmaltreated: $M = 6$ years 5 months, $SD = 1.81$), $t(297) = 1.30$, $p = .165$; gender composition, $\chi^2(1) = 0.103$, $p = .748$; and productive vocabulary as assessed by the Woodcock–Johnson (maltreated: $M = 23.92$, $SD = 4.0$, mental age equivalent = 5 years 7 months; nonmaltreated: $M = 24.15$, $SD = 4.0$, mental age equivalent = 5 years 8 months), $t(285) = 0.49$, $p = .627$. The maltreated and nonmaltreated children were also comparable in percentages of Latinos (maltreated: 64%, nonmaltreated: 59%), $\chi^2(1) = 0.84$, $p = .360$, and Caucasians (maltreated: 8%, nonmaltreated: 4%), $\chi^2(1) = 1.84$, $p = .175$. However, a somewhat lower percentage of maltreated children were African American (maltreated: 22%, nonmaltreated: 33%), $\chi^2(1) = 4.26$, $p = .039$.

Disclosure Endorsement

We first analyzed whether children endorsed disclosure or nondisclosure and then analyzed their explanations for their endorsement. Children received a 1 for a *yes* response (endorsing disclosure) and a 0 for a *no* response (endorsing nondisclosure). Only 1% of responses were not codable as *yes* or *no* (*don't know*, unscorable). These are not considered further.

Preliminary analyses revealed no main effects or two-way interactions because of ethnicity, gender, or verbal ability, and these factors were excluded from subsequent analyses. Moreover, children did not discriminate between mother and father instigators or mother and father recipients. We thus collapsed across mothers and fathers to create mean scores for each child for parent instigator and parent recipient.

Omnibus analysis. The means across maltreatment status and age for instigator identity, recipient identity, and instigator–recipient scenario are presented in Tables 1 and 2. Preliminary analyses showed that the effects of the should–would ("should" the child tell vs. "would" the child tell) and transgression severity ("really bad" vs. "just a

Table 1

Proportions (SD) of Disclosure Endorsement to Different Instigators and Recipients, by Age and Maltreatment Status

	Instigator conditions		Recipient conditions			
	Parent	Stranger	Police	Parent	Teacher	Grand mean
4–5 years						
Maltreated	.53 (0.44)	.55 (0.42)	.56 (0.44)	.56 (0.44)	.54 (0.44)	.54 (0.43)
Nonmaltreated	.71 (0.34)	.77 (0.33)	.73 (0.35)	.76 (0.33)	.68 (0.35)	.73 (0.32)
6–7 years						
Maltreated	.54 (0.39)	.64 (0.41)	.63 (0.41)	.61 (0.41)	.49 (0.43)	.58 (0.38)
Nonmaltreated	.49 (0.39)	.62 (0.36)	.61 (0.39)	.63 (0.37)	.36 (0.39)	.54 (0.34)
8–9 years						
Maltreated	.60 (0.30)	.79 (0.26)	.74 (0.27)	.73 (0.30)	.53 (0.36)	.67 (0.26)
Nonmaltreated	.42 (0.34)	.77 (0.25)	.54 (0.30)	.67 (0.28)	.42 (0.32)	.56 (0.26)

Note. "Disclosure endorsement" refers to answering "yes" to "should s/he tell" or "would s/he tell."

Table 2

Proportions (SD) of Disclosure Endorsement to Different Instigator–Recipient Pairs, by Age and Maltreatment Status

	Instigator–recipient pairs					
	Parent–police	Stranger–police	Parent–parent	Stranger–parent	Parent–teacher	Stranger–teacher
4–5 years						
Maltreated	.56 (0.45)	.56 (0.46)	.54 (0.47)	.54 (0.44)	.49 (0.44)	.57 (0.46)
Nonmaltreated	.71 (0.39)	.77 (0.36)	.75 (0.38)	.80 (0.34)	.66 (0.38)	.69 (0.36)
6–7 years						
Maltreated	.60 (0.42)	.67 (0.44)	.57 (0.43)	.66 (0.43)	.46 (0.44)	.56 (0.47)
Nonmaltreated	.53 (0.43)	.76 (0.40)	.57 (0.43)	.68 (0.40)	.36 (0.43)	.37 (0.47)
8–9 years						
Maltreated	.66 (0.36)	.91 (0.24)	.70 (0.36)	.76 (0.34)	.44 (0.42)	.72 (0.38)
Nonmaltreated	.38 (0.40)	.86 (0.30)	.56 (0.38)	.80 (0.28)	.32 (0.37)	.62 (0.40)

Note. "Disclosure endorsement" refers to answering "yes" to "should s/he tell" or "would s/he tell."

little bad'') manipulations were limited to specific age groups, and those effects are discussed next. To test the study's main hypotheses, we conducted a 2 (maltreatment) × 3 (age: 4–5, 6–7, 8–9) × 2 (instigator: parent, stranger) × 3 (recipient: parent, police officer, teacher) mixed model analysis of variance (ANOVA).

Two significant main effects emerged: instigator, $F(1, 292) = 79.77$, $p < .001$, partial $\eta^2 = .22$, and recipient, $F(2, 292) = 58.86$, $p < .001$, partial $\eta^2 = .22$. Children endorsed lower rates of disclosure against parents ($M = 0.55$, $SD = 0.37$) than strangers ($M = 0.69$, $SD = 0.35$), and endorsed lower rates of disclosure to teachers ($M = 0.50$, $SD = 0.39$) than parents ($M = 0.67$, $SD = 0.36$) or the police ($M = 0.64$, $SD = 0.37$).

The effects of maltreatment and age were not significant, but several significant interactions emerged involving these variables. Specifically, maltreatment interacted with age, $F(2, 292) = 5.65$, $p = .004$, partial $\eta^2 = .04$; instigator, $F(1, 292) = 4.15$, $p = .043$, partial $\eta^2 = .01$; and recipient, $F(2, 292) = 5.17$, $p = .006$, partial $\eta^2 = .02$. Similarly, age interacted with instigator, $F(2, 292) = 23.21$, $p < .001$, partial $\eta^2 = .14$, and recipient, $F(4, 292) = 6.51$, $p < .001$, partial $\eta^2 = .04$. The two-way instigator and recipient interaction, reflecting differences across the six scenarios, was significant, $F(4, 584) = 7.69$, $p = .001$, partial $\eta^2 = .03$, as were the Instigator × Recipient × Maltreatment interaction, $F(2, 584) = 10.05$, $p < .001$, partial $\eta^2 = .03$, and the Instigator × Recipient × Age interaction, $F(4, 584) = 5.72$, $p < .001$, partial $\eta^2 = .04$.

To interpret these interactions, we examined age effects for maltreated and nonmaltreated children across the different instigator–recipient pairs, conducting one-way ANOVAs with age as the between-subjects factor. To reduce the likelihood of Type I errors, we used the Bonferroni correction (for 12 analyses) and set the critical p value at .004. For the maltreated children, only one type of story showed an age effect: Older children were *more* likely with age to endorse disclosure in the stranger–police scenarios, $F(2, 150) = 11.43$, $p < .001$, partial $\eta^2 = .13$. For the nonmaltreated children, older children were *less* likely to endorse disclosure in the parent–parent scenarios, $F(2, 143) = 9.64$, $p < .001$, partial $\eta^2 = .12$, and the parent–police scenarios, $F(2, 143) = 9.73$, $p < .001$, partial $\eta^2 = .12$. The nonmaltreated children also exhibited age effects for the parent–teacher scenarios, $F(2, 143) = 12.29$, $p < .001$, partial $\eta^2 = .15$, and stranger–teacher scenarios, $F(2, 143) = 7.59$, $p = .001$, partial $\eta^2 = .10$, which were attributable to decreases in disclosure endorsement to teachers by 6–7 years of age (a trend that was reversed in the stranger–teacher scenarios among the oldest children).

To summarize, children endorsed less disclosure against parent instigators than against stranger instigators, consistent with our predictions. Unexpectedly, children endorsed less disclosure to teachers than to either parents or the police. Neither maltreatment nor age had straightforward effects. Rather, maltreated children changed little in their endorsement of disclosure with age, except that they endorsed more disclosure with age in the stranger–police scenarios. Nonmaltreated children showed some evidence of decreased endorsement of disclosure with age, particularly when parents were instigators. To further understand the different patterns because of maltreatment status and age, we separately analyzed each age group.

Within age effects: 4- to 5-year-olds. Preliminary analyses revealed that the should–would manipulation (''should'' the child tell vs. ''would'' the child tell) affected children's responses at this age. A 2 (maltreatment) × 2 (instigator) × 3 (recipient) × 2 (should–would) ANOVA revealed significant effects of maltreatment, $F(1, 92) = 6.72$, $p = .011$, partial $\eta^2 = .07$; instigator, $F(1, 92) = 4.48$, $p = .04$, partial $\eta^2 = .05$; and recipient, $F(1, 184) = 4.61$, $p = .011$, partial $\eta^2 = .05$, and an interaction between should–would and instigator, $F(1, 92) = 11.42$, $p < .001$, partial $\eta^2 = .11$.

The main effect because of maltreatment was attributable to the fact that maltreated children ($M = 0.54$, $SD = 0.43$) endorsed lower rates of disclosure than nonmaltreated children ($M = 0.73$, $SD = 0.32$), $t(95) = 2.58$, $p = .011$. The main effect of instigator was because of children's greater endorsement of disclosure against strangers ($M = 0.66$, $SD = 0.39$) than parents ($M = 0.63$, $SD = 0.40$). This effect was qualified, however, by the interaction between instigator and should–would. It was only when asked what the child ''would'' do that children endorsed more disclosure against strangers ($M = 0.66$, $SD = 0.37$) than against parents ($M = 0.55$, $SD = 0.40$), $t(45) = 3.35$, $p = .002$. When asked what the child ''should'' do, children were equally likely to endorse disclosure against strangers ($M = 0.67$, $SD = 0.40$) as against parents ($M = 0.69$, $SD = 0.39$), $t(50) = 1.30$, $p = .201$. The main effect because of recipient was attributable to lower rates of endorsed disclosure to teachers ($M = 0.60$, $SD = 0.40$) than parents ($M = 0.66$, $SD = 0.40$), $t(96) = 3.36$, $p = .001$, or the police ($M = 0.65$, $SD = 0.40$), $t(96) = 2.30$, $p = .024$.

Together, these results suggest that, among the youngest children, nonmaltreated children endorsed more disclosure than maltreated children. Children endorsed more disclosure against stranger instigators than against parent instigators when asked what the story child "would" do but not when asked what the story child "should" do. Finally, children endorsed less disclosure to teachers than to police or parents.

6- to 7-year-olds. A 2 (maltreatment) × 2 (instigator) × 3 (recipient) ANOVA revealed significant main effects because of instigator, $F(1, 91) = 14.17$, $p < .001$, partial $\eta^2 = .13$; recipient, $F(1, 182) = 29.06$, $p < .001$, partial $\eta^2 = .24$; and an interaction between instigator and recipient, $F(2, 182) = 2.99$, $p = .053$, partial $\eta^2 = .03$. Maltreatment did not exhibit a main effect but interacted with recipient, $F(2, 186) = 4.53$, $p = .012$, partial $\eta^2 = .05$, and there was a three-way interaction among maltreatment, instigator, and recipient, $F(2, 186) = 5.23$, $p = .006$, partial $\eta^2 = .05$. To understand the interactions, we analyzed the different pairings of instigator and recipient separately (Table 2).

We had predicted that maltreated children would be less inclined than nonmaltreated children to protect parents as instigators or prefer parents as recipients. Therefore, we conducted a series of within-subject *t* tests in which we focused on comparisons between parents and nonparents. We first compared instigators (parents vs. strangers) for each type of recipient (parent, police, or teacher), and then compared recipients (parents vs. police) for each type of instigator (parent or stranger). Given the number of analyses (10), we adjusted the critical *p* value (.005). When the recipient was a police officer, only nonmaltreated children distinguished between parent and stranger transgressors, endorsing less disclosure against parents, $t(41) = 4.16$, $p < .001$. None of the other comparisons between parent and stranger instigators were significant (parent vs. strangers-teachers or parent vs. strangers-parents). Neither group of children exhibited a preference for parent over police recipients.

These results reveal that the 6- to 7-year-olds were, in some respects, similar to the younger children, endorsing more disclosure against strangers than against parents. However, there were no longer overall differences in endorsement of disclosure between nonmaltreated and maltreated children. Instead, nonmaltreated children began to discriminate more than maltreated children between parent and stranger instigators when talking to the police.

8- to 9-year-olds. Preliminary analyses revealed that the severity of the transgression ("really bad" vs. "just a little bad") affected children's responses at this age. Therefore, we conducted a 2 (maltreatment) × 2 (instigator) × 3 (recipient) × 2 (transgression severity) ANOVA. There were significant main effects of instigator, $F(1, 103) = 79.18$, $p < .001$, partial $\eta^2 = .44$, and recipient, $F(2, 206) = 26.33$, $p < .001$, partial $\eta^2 = .20$, and an interaction between instigator and recipient $F(2, 206) = 12.26$, $p < .001$, partial $\eta^2 = .11$. There were also main effects because of the severity of the transgression, $F(1, 103) = 18.06$, $p < .001$, partial $\eta^2 = .15$, and maltreatment status, $F(1, 103) = 5.32$, $p = .023$, partial $\eta^2 = .05$. The main effect for transgression severity was attributed to greater disclosure endorsement in response to "really bad" ($M = 0.66$, $SD = 0.27$) than "just a little bad" ($M = 0.57$, $SD = 0.29$) transgressions.

The main effect for maltreatment status reflected greater endorsement of disclosure among the maltreated children (Table 1). However, there was a three-way interaction among maltreatment, instigator, and recipient, $F(2, 206) = 4.07$, $p = .018$, partial $\eta^2 = .04$. Similar to the previous age group, we sought to understand the interactions by analyzing the different pairings of instigator and recipient separately, focusing on differences between parents and strangers as instigators and parents and police as recipients (Table 2). Again, because of multiple tests (10), the critical *p* value was set at .005.

When talking to the police, both maltreated and nonmaltreated children endorsed more disclosure against strangers than against parents—maltreated, $t(54) = 4.89$, $p < .001$, and nonmaltreated, $t(51) = 7.59$, $p < .001$—although the differences were larger for the nonmaltreated children. However, when talking to parents, only the nonmaltreated children were less likely to disclose against parents than against strangers: maltreated, $t(54) = 1.35$, $p = .183$, and nonmaltreated, $t(51) = 4.67$, $p < .001$. When deciding whether to disclose parental transgressions, only the nonmaltreated children preferred parent recipients to police recipients, $t(51) = 7.59$, $p < .001$. In contrast, when deciding whether to disclose stranger transgressions, only the maltreated children preferred police recipients over parent recipients: maltreated, $t(54) = 2.93$, $p = .005$, and nonmaltreated, $t(51) = 1.41$, $p = .17$.

To summarize the performance of the oldest children, they were the first to make distinctions on the basis of transgression severity ("just a little bad" vs. "really bad"). The maltreated and nonmaltreated children exhibited differences in their reactions to different instigator–recipient pairs, with the maltreated children sometimes exhibiting less deference to parental instigators and less preference for parental recipients than nonmaltreated children. Although both groups of children were more likely to endorse disclosure against strangers than against parents when talking to the police, the maltreated children were just as likely to endorse disclosure against parents as against strangers when another parent was the recipient. Finally, it is at 8–9 years of age when children first exhibited preferences for certain recipients given certain instigators. The nonmaltreated children preferred parent recipients when reporting on parent instigators, and the maltreated children preferred police recipients when reporting on stranger instigators. Hence, there was some evidence that the maltreated children were less deferential than nonmaltreated children to parental instigators and less likely to prefer parents as recipients.

Explanations

After children endorsed disclosure or nondisclosure, they were asked "why." We reliably coded children's explanations for whether they referred to various components of the scenarios, with the goal of identifying potentially significant concepts that could explain whether they selected disclosure or nondisclosure. Four primary categories were identified and children's responses could fall into more than one category: (a) instigator (e.g., "[Instigator] did not want him to tell"), (b) recipient (e.g., "[Recipient] will be able to solve it"), (c) transgression (e.g., "It's nothing important"), and (d) command (e.g., "[Instigator] said, 'Don't tell'"). We also coded for any reference to promising or secrecy, or to truth and lies to determine if children appeared to equate the scenarios with situations in which a child promises to keep a secret or lies. There were four coders (one gold standard coder and three research assistants). Based on 200 explanations, interrater agreement between the gold standard coder and the research assistants was 96%, with all Kappas exceeding .88. Discrepancies during the reliability coding were resolved through discussion.

We expected children's explanations to vary depending on scenario type and whether the child endorsed disclosure (i.e., answered "yes" or "no"). Because some children provided a mix of yes and no responses about disclosure to the different scenarios, whereas other children consistently responded "yes" or "no," the responses could not be appropriately analyzed either as between subjects or repeated measures; we therefore utilized a modified version of the t test that allows adjustment for the extent to which some responses are correlated within subjects (Looney & Jones, 2003).

We anticipated that whether children mentioned the instigator would depend on whether a stranger or a parent was the instigator and whether the child endorsed disclosure. We hypothesized that children who endorsed disclosure would reference the instigator more often in the stranger than parent instigator stories (because the identity of the instigator would justify disclosure in the stranger stories). This prediction was supported (corrected: $z = 4.04$, $p < .001$; stranger: $M = 0.67$, $SD = 0.34$; parent: $M = 0.60$, $SD = 0.35$). We expected that children who endorsed nondisclosure should reference the instigator more often in the parent than stranger instigator stories (because the identity of the instigator would justify nondisclosure in the parent stories). This prediction was also supported (corrected: $z = 2.34$, $p = .001$; stranger instigator: $M = 0.51$, $SD = 0.44$; parent: $M = 0.58$, $SD = 0.39$).

With respect to the mention of transgression, we expected that, if children endorsed disclosure, they would mention the transgression more often when the transgression was "really bad" than when it was "just a little bad." This prediction was not supported (corrected: $z = .54$, $p = .29$; really bad: $M = 0.33$, $SD = 0.36$; little bad: $M = 0.32$, $SD = 0.34$). If children endorsed nondisclosure, we expected the opposite; they would mention the transgression more often when the transgression was "just a little bad" than when it was "really bad." This prediction was also not supported (corrected: $z = .40$, $p = .66$; really bad: $M = 0.10$, $SD = 0.25$; little bad: $M = 0.10$, $SD = 0.22$).

Finally, with reference to the command category (in which children referred to the admonishment "don't tell"), we expected that this reference would occur more often when children endorsed nondisclosure than when they endorsed disclosure. This prediction was not supported, although the difference approached statistical significance (corrected: $z = 1.38$, $p = .08$; yes responses: $M = 0.08$, $SD = 0.22$; no responses: $M = 0.37$, $SD = 0.35$). Children only very rarely referred to promises or secrets (yes responses: $M = 0.02$, $SD = 0.09$; no responses: $M = 0.06$, $SD = 0.20$) or to truth and lies (yes responses: $M = 0.02$, $SD = 0.08$; no responses: $M = 0.01$, $SD = 0.07$).

The unexpected finding that children were less likely to endorse disclosure to teachers (noted earlier) led us to examine whether children who endorsed disclosure referenced the recipient more often in their explanations in the parent and police recipient stories than in the teacher recipient stories (because the identity of the recipient would justify disclosure in the parent and police stories), and whether children who endorsed nondisclosure would reference the recipient more often in the teacher recipient stories. No differences emerged in how often children who endorsed disclosure mentioned the recipient (corrected: $z = 1.17$, $p = .12$; parent/police: $M = 0.45$, $SD = 0.39$; teacher: $M = 0.47$, $SD = 0.41$). However, children who endorsed nondisclosure referred more often to the teacher recipient than to the parent or police recipient (corrected: $z = 3.01$, $p = .001$; parent/police: $M = 0.31$, $SD = 0.40$; teacher: $M = 0.41$, $SD = 0.39$).

In sum, some of our predictions with respect to explanations were borne out, especially when children endorsed nondisclosure (i.e., answered "no" to "should–would [the story child] tell"). Children distinguished between stranger and parent instigators in their reasoning about disclosure, and when endorsing nondisclosure, children distinguished between teachers and other recipients. Children only very rarely referred to promises or truth and lies, providing little evidence that they interpreted nondisclosure as a lie or the violation of a promise.

Discussion

This is the first study to examine maltreated and nonmaltreated children's attitudes about disclosure of adult transgressions varying the relationship of child to instigator (a transgressor who commands secrecy) and to recipient (an adult to whom the child might disclose). We examined a wide age range (4–9 years of age) and compared maltreated children to demographically similar nonmaltreated children. Our findings revealed that children's judgments were related to age and maltreatment in important ways. Moreover, the identity of the instigator and recipient affected children's willingness to endorse disclosure of adult transgressions. Each is discussed in turn.

Age and Maltreatment

Our inclusion of a wide age range and children with diverse maltreatment histories enabled us to consider developmental and experiential contributions to children's attitudes about transgression disclosure. We anticipated that children would be more inclined to endorse disclosure with age. Research on children's attitudes about obedience to authority suggests that unless a commanded act is unambiguously immoral, younger children tend to be more deferential than older children to authority (Laupa, 1994; Laupa, et al., 1995; Piaget, 1932). We thought that young children in particular would respond to the stories as if the instigator's admonition (don't tell) was an authoritative command, reducing their likelihood of endorsing disclosure. With respect to maltreatment, we suspected that maltreated children would anticipate more negative instigator reactions (and less supportive recipient reactions) and therefore endorse less disclosure overall. Across different instigators and recipients, we expected that maltreated children would be less inclined than nonmaltreated children to protect parents as instigators or prefer parents as recipients.

The results revealed that it was necessary to consider the effects of age and maltreatment together. The youngest maltreated children did in fact endorse less disclosure than the nonmaltreated children, but this difference disappeared among the 6- to 7-year-olds and was reversed among the 8- to 9-year-olds. Examining age differences within the two groups revealed that with respect to talking to the police, the maltreated and nonmaltreated children exhibited opposite age trends. The maltreated children were more inclined with age to disclose stranger transgressions. The nonmaltreated children, however, grew increasingly protective of parental instigators. As a result, older maltreated children were sometimes less likely than nonmaltreated children to distinguish between parents and strangers when deciding whether or not to endorse disclosure.

The finding that nonmaltreated children endorsed less disclosure against parent instigators with age appears most similar to findings that children over this age range endorse less disclosure when peers commit transgressions (Piaget, 1932; Watson & Valtin, 1997). Our findings suggest that attitudes about secrecy apply to child–parent relationships as well as child–peer relationships. The maltreatment differences may mirror findings that maltreated children have less faith in the supportiveness of parents than nonmaltreated children (Macfie et al., 1999; Shields et al., 2001; Shipman & Zeman, 2001; Shipman et al., 2003; Toth et al., 2000), which might explain why older maltreated children are not more protective of parent instigators than younger children.

Instigator Identity

Consistent with our prediction, children were less likely to endorse disclosure when the instigator was a parent rather than a stranger, and this pattern was confirmed by children's explanations. Even the 4- to 5-year-olds distinguished between parents and strangers, although these young children only endorsed less disclosure against a parent instigator when asked what a story child "would" rather than "should" do. This suggests that the youngest children were aware that disclosure is *less* likely against parents, whereas by 6–7 years of age, children both respond that one "would" and "should" disclose less against parents. Less endorsement of disclosure against parents is consistent with research examining children's actual disclosure of transgressions, both in experimental work on children's lies (Tye et al., 1999) and in observational research examining patterns of sexual abuse disclosure (Hanson et al., 1999; London et al., 2008; Lyon, 2009). Ours is the first study to demonstrate that these patterns are also reflected in children's attitudes.

Recipient Identity

The identity of the individual to whom the story child may disclose also affected children's responses, although in unexpected ways. Children endorsed less disclosure to teachers, and when they endorsed nondisclosure, they referred to the recipient more often in teacher recipient vignettes than when the recipient was a parent or the police. We suspect this is attributable to children's lacking understanding of the potential role of teachers in intervention; children may not be aware, for example, that teachers are mandated reporters of child abuse.

We predicted that children would generally be more inclined to endorse disclosure when the parent was a recipient. This prediction was not supported. Rather, preference for parental recipients emerged only in some scenarios and only among the oldest children. Moreover, this preference was limited to the nonmaltreated children, whereas maltreated children sometimes exhibited a preference for police recipients. The differences might reflect a relative lack of trust in parents among maltreated children.

Severity of the Transgression

Very few significant effects emerged regarding transgression severity. In fact, children did not take the severity of the transgression into account until they were 8–9 years of age. On the one hand, this may reflect the relative importance of instigators and recipients in children's judgments about transgression secrecy. On the other hand, it may be the case that children are influenced by the types of transgressions but that they require specific examples of those transgressions to make their judgments. In the future, it will be important to examine, using specific transgressions that vary in severity, whether and at what age children take severity into account when rendering decisions about disclosure.

Limitations

Although the findings reported here are novel and provide new insight into the combined effects of instigator and recipient relationships on maltreated and nonmaltreated children's emerging attitudes about transgression disclosure, limitations must also be acknowledged.

As a result of our interest in examining the relation between maltreatment and children's attitudes and because the research was located in a large urban area, our maltreated sample and comparable nonmaltreated sample were predominantly low income and of ethnic minority status. In contrast, prior research on children's attitudes about secrecy has tended to examine the attitudes of Caucasian children from middle- and upper-middle-class homes. Given the fact that this is the first study to examine children's attitudes about disclosure of adult transgressions, research is needed to examine whether the developmental differences observed here replicate more generally.

It might be the case, for example, that children with different backgrounds develop different attitudes about authority figures (e.g., police and teachers), and this could have influenced their judgments in the scenarios in which authority figures were the potential recipients. Indeed, there is considerable evidence that adolescents of ethnic minority status, particularly in low-income communities, tend to endorse negative attitudes about the police (Hurst & Frank, 2000; Sullivan, Dunham, & Alpert, 1987; Taylor, Turner, Esbensen, & Winfree, 2001) and that adolescents who endorse less positive attitudes toward the police also tend to endorse less positive attitudes toward other adults, including teachers and parents (Flexon, Lurigio, & Greenleaf, 2009; Nihart, Lersch, Sellers, & Mieczkowski, 2005).

However, there are hints in this research that participants in our study were less likely to be influenced by such attitudes. For one, we examined attitudes among elementary-aged children, a population that is not represented in the authority attitudinal research. The little developmental research that has been conducted suggests that the negative attitudes emerge as children move from preadolescence to adolescence. This may be because it is during adolescence when young people begin to attract the notice of police as potential suspects (Fagan & Tyler, 2005). Second, research on adolescents' attitudes tends to find that Latinos express less negative attitudes than African Americans (Sullivan et al., 1987; Taylor et al., 2001), and yet we found no ethnic differences in our sample. The dynamics between children's willingness to disclose to the police and other authorities in relation to children's attitudes about these professionals is clearly of interest in its own right and worthy of continued research.

Another limitation is that the scenarios did not specify what the adult instigator did; rather, we referred to the actions as "just a little bad" or "really bad." Scenarios were designed to leave the acts unspecified; this was because our primary interest was how children reasoned about disclosure given different instigators and recipients. We do not know whether children imagined specific content, and if so, what sort of actions they imagined. For example, children may have imagined moral acts of varying severity or as violations of social conventions. Of importance, however, several aspects of the task suggested that the transgressions reflected immoral acts. For one, the instigator gave the admonition "don't tell" to the child protagonist in every vignette and the police were included as potential recipients in a quarter of the vignettes (and, by the early school years, a majority of both maltreated and nonmaltreated children have at least a rudimentary understanding of the function of police; Cooper, Quas, Wallin, & Lyon, 2010). Also, prior work suggests that describing acts as "really bad," as was done in the current study, strongly connotes immoral behavior (Smetana et al., 1984). Future research can profitably explore whether children's attitudes about disclosure are in fact affected by the specific type of transgression. For example, we would predict that children will endorse disclosure of a peer's moral violation much more readily than they will endorse disclosure of a peer's violation of social conventions.

On a related note, differences in children's attitudes might have been attributed to children making different assumptions about the actions that different instigators performed. This seems unlikely, because transgression severity had minimal effects, both in terms of children's overall disclosure endorsements and their explanations for their responses. For example, children's explanations often included mention of identities of instigators and recipients without any mention of the transgression. If children were relatively insensitive to our explicit references to severity, one would not expect children to infer different actions that were described in an identical fashion across scenarios. Nevertheless, to address this concern future research could provide children more specific information about the transgressions in issue or inquire into children's assumptions about the transgressions that occurred.

Whether children's reasoning about the vignettes truly captures their reasoning with respect to their own decision making is of course subject to question, a problem that Piaget (1932) long ago recognized when questioning children about their moral judgments. One way to test this is to examine whether individual differences in children's reasoning predict differences in disclosure behavior. Research examining relations between children's attitudes about the morality of lying and honesty has tended to find low correlations (London & Nunez, 2002; Lyon, et al., 2008; Pipe & Wilson, 1994; Talwar et al., 2002). However, no research has examined the possible relation between children's attitudes about secrecy and their failure to volunteer information, as opposed to their attitudes about and proclivity toward lying. Examining individual differences in children's attitudes about transgression secrecy is also of interest in its own right; one could move toward understanding the underlying reasons for children's distinctions by searching for relations between children's attitudes about disclosure and other attitudes, such as the aforementioned trust in adults.

Future Directions

In examining children's attitudes about transgression disclosure, this study focused exclusively on adult instigators, given their obvious relevance to child maltreatment. But because the results suggest that nonmaltreated children's increasing protectiveness of their parents may parallel their attitudes about peers, future research should include both adult and peer instigators. Similarly, our assumptions that the patterns of disclosure endorsement are related to trust in adults, with trust in an instigator leading to less disclosure endorsement against that instigator and with trust in a recipient leading to greater disclosure endorsement to that confidant, are worthy of direct test. Although researchers have not explored the relation between trust and secrecy with respect to young children's relationship to their parents, researchers examining developing conceptions of friendship have noted such a connection (Rotenberg, 1991). Future work could explicitly test for links among children's expectations of adult supportiveness, adult reactions to disclosure, and their disclosure endorsement, and determine whether these links vary between maltreated and nonmaltreated children.

Future research might also assess the influence of other factors on children's attitudes about transgression disclosure. In this study, the instigator in the vignettes instructed the child not to tell. Whether this influenced children's judgments is unclear. If children based their judgments on their trust in the different adults involved, rather than fear of retribution, then the admonition may not have had an effect. Conversely, stronger language by the instigator might affect children's judgments. For example, explicit threats of punishment (or, in the case of parents, loss of love) might have an effect, as might explicit requests to keep the transgression a secret, or elicitation of a promise by the story child not to disclose. At the same time, it would be interesting to assess children's attitudes about various approaches that recipients might take to encourage disclosure.

Finally, future research with maltreated children might attempt to tease apart whether specific aspects of maltreatment influence children's judgments about transgression disclosure, including, for example, the effects of severity or type of maltreatment, children's relationship to the perpetrator, and the possibly confounding effects of state intervention. One might, for example, interpret the finding in our study that older maltreated children endorse more disclosure than nonmaltreated children as reassuring, because it suggests that children who experience maltreatment do not think that transgressions should be kept a secret. However, the maltreated children in this study were of necessity those children who had been identified as such by the state, and were extreme cases of maltreatment because the children had been removed from their parents' custody. It may be that children who suffer from isolated types of maltreatment (such as sexual abuse in an otherwise non-neglectful and nonabusive home) maintain positive attitudes toward transgression secrecy.

In conclusion, this study provides a first step toward an understanding of how children reason about disclosing the wrongdoing of adults, particularly those close to them. Theoretically, the findings contribute to our understanding of children's developing attitudes about the effect of social relationships on secrecy and the effects that adverse experiences may have on those attitudes. Practically, the results suggest that children's deference to parents may be a double-edged sword. Protectiveness of parents is surely a virtuous trait. In extreme circumstances, however, it may hinder society's ability to help children whose trust is misplaced.

References

Anderson, J., Martin, J., Mullen, P., Romans, S., & Herbison, P. (1993). Prevalence of childhood sexual abuse experiences in a community sample of women. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*, 911–919.

Bok, S. (1982). *Secrecy*. New York: Pantheon.

Buhrmester, D., & Prager, K. (1995). Patterns and functions of self-disclosure during childhood and adolescence. In K. J. Rotenberg (Ed.), *Disclosure processes in children and adolescents* (pp. 10–56). New York: Cambridge University Press.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development*, *63*, 129–137.

Bussey, K. (1993, March). *Factors influencing children's disclosure of witnessed events*. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development*, *70*, 1338–1347.

Bussey, K., & Grimbeek, E. J. (1995). Disclosure processes in children and adolescents. In K. J. Rotenberg (Ed.), *Disclosure processes in children and adolescents* (pp. 166–203). New York: Cambridge University Press.

Cooper, A., Quas, J. A., Wallin, A. R., & Lyon, T. D. (2010). Maltreated and nonmaltreated children's knowledge of the juvenile dependency court system. *Child Maltreatment*, *15*, 255–260.

Damon, W. (1977). *The social world of the child*. San Francisco: Jossey-Bass.

Fagan, J., & Tyler, T. R. (2005). Legal socialisation of children and adolescents. *Social Justice Research*, *18*, 217–241.

Fleming, J. M. (1997). Prevalence of childhood sexual abuse in a community sample of Australian women. *Medical Journal of Australia*, *166*, 65–68.

Flexon, J. L., Lurigio, A. J., & Greenleaf, R. G. (2009). Exploring the dimensions of trust in the police among

Chicago juveniles. *Journal of Criminal Justice*, *37*, 180–189.

Hanson, R. F., Resnick, H. S., Saunders, B. E., Kilpatrick, D. G., & Best, C. (1999). Factors related to the reporting of childhood rape. *Child Abuse & Neglect*, *23*, 559–569.

Hershkowitz, I., Lanes, O., & Lamb, M. E. (2007). Exploring the disclosure of child sexual abuse with alleged victims and their parents. *Child Abuse & Neglect*, *31*, 111–123.

Huffman, M. L., Warren, A. R., & Larson, S. M. (1999). Discussing truth and lies in interviews with children: Whether, why, and how? *Applied Developmental Science*, *1*, 6–15.

Hunter, F. T., & Youniss, J. (1982). Changes in functions of three relations during adolescence. *Developmental Psychology*, *18*, 306–311.

Hurst, Y. G., & Frank, J. (2000). How kids view cops: The nature of juvenile attitudes toward the police. *Journal of Criminal Justice*, *28*, 189–202.

Kogan, S. M. (2004). Disclosing unwanted sexual experiences: Results from a national sample of adolescent women. *Child Abuse & Neglect*, *24*, 147–165.

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened. Structured investigative interviews of child victims and witnesses*. West Sussex, UK: Wiley.

Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.

Laupa, M. (1994). Who's in charge? Preschool children's concepts of authority. *Early Childhood Research Quarterly*, *9*, 1–17.

Laupa, M., & Turiel, E. (1986). Children's conceptions of adult and peer authority. *Child Development*, *57*, 405–412.

Laupa, M., Turiel, E., & Cowan, P. (1995). Obedience to authority in children and adults. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 131–165). New York: Cambridge University Press.

Lawson, L., & Chaffin, M. (1992). False negatives in sexual abuse disclosure interviews: Incidence and influence of caretaker's belief in abuse in cases of accidental abuse discovery by diagnosis of STD. *Journal of Interpersonal Violence*, *7*, 532–542.

London, K., Bruck, M., Wright, D. B., & Ceci, S. J. (2008). Review of the contemporary literature on how children report sexual abuse to others: Findings, methodological issues, and implications for forensic interviewers. *Memory*, *16*, 29–47.

London, K., & Nunez, N. (2002). Examining the efficacy of truth/lie discussions in predicting and increasing the veracity of children's reports. *Journal of Experimental Child Psychology*, *83*, 131–147.

Looney, S. W., & Jones, P. W. (2003). A method for comparing two normal means using combined samples of correlated and uncorrelated data. *Statistics in Medicine*, 22, 1601–1610.

Lyon, T. D. (2002). Scientific support for expert testimony on child sexual abuse accommodation. In J. R. Conte (Ed.), *Critical issues in child sexual abuse* (pp. 107–138). Thousand Oaks, CA: Sage.

Lyon, T. D. (2009). Abuse disclosure: What adults can tell. In B. L. Bottoms, G. S. Goodman, & C. J. Najdowski (Eds.), *Child victims, child offenders: Psychology and law.* (pp. 19–35). New York: Guilford.

Lyon, T. D., Carrick, N., & Quas, J. A. (2010). Young children's competency to take the oath: Effects of task, maltreatment, and age. *Law & Human Behavior*, *34*, 141–149.

Lyon, T. D., & Dorado, J. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect*, *32*, 738–748.

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, *79*, 914–929.

Lyon, T. D., & Saywitz, K. J. (1999). Young maltreated children's competence to take the oath. *Applied Developmental Science*, *3*, 16–27.

Macfie, J., Toth, S. L., Rogosch, F. A., Robinson, J., Emde, R. N., & Cicchetti, D. (1999). Effect of maltreatment on preschoolers' narrative representations of responses to relieve distress and of role reversal. *Developmental Psychology*, *35*, 460–465.

Malloy, L. C., Lyon, T. D., & Quas, J. A. (2007). Filial dependency and recantation of child sexual abuse allegations. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*, 162–170.

Mullen, P. E., Martin, J. L., Anderson, J. C., Romans, S. E., & Herbison, G. P. (1993). Childhood sexual abuse and mental health in adult life. *British Journal of Psychiatry*, *163*, 721–732.

Nihart, T., Lersch, K., Sellers, C., & Mieczkowski, T. (2005). Kids, cops, parents and teachers: Exploring juvenile attitudes toward authority figures. *Western Criminology Review*, *6*, 79–88.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect*, *24*, 733–752.

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in the ideas about lying. *Child Development*, *54*, 1529–1535.

Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt Brace Jovanovich.

Pipe, M., & Goodman, G. S. (1991). Elements of secrecy: Implications for children's testimony. *Behavioral Sciences and the Law*, *9*, 33–41.

Pipe, M., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology*, *30*, 515–525.

Rotenberg, K. J. (1991). The trust-value basis of children's friendship. In K. J. Rotenberg (Ed.), *Children's interper-*

sonal trust: Sensitivity, lying, deception, and promise violations (pp. 160–172). New York: Springer-Verlag.

Rotenberg, K. (1995). Moral development and children's differential disclosure to adults versus peers. In K. J. Rotenberg (Ed.), *Disclosure processes in children and adolescents* (pp. 135–147). New York: Cambridge University Press.

Russell, D. E. H. (1983). The incidence and prevalence of intrafamilial and extrafamilial sexual abuse of female children. *Child Abuse & Neglect*, *7*, 133–146.

Shields, A., Ryan, R. M., & Cicchetti, D. (2001). Narrative representations of caregivers and emotional dysregulation as predictors of maltreated children's rejection by peers. *Developmental Psychology*, *37*, 321–337.

Shipman, K. L., & Zeman, J. (2001). Socialization of children's emotion regulation in mother-child dyads: A developmental psychopathology perspective. *Development and Psychopathology*, *13*, 317–336.

Shipman, K., Zeman, J., Fitzgerald, M., & Swisher, L. M. (2003). Regulating emotion in parent-child and peer relationships: A comparison of sexually maltreated and nonmaltreated girls. *Child Maltreatment*, *8*, 163–172.

Smetana, J. G., Daddis, C., Toth, S. L., Cicchetti, D., Bruce, J., & Kane, P. (1999). Effects of provocation on maltreated and nonmaltreated preschoolers' understanding of moral transgressions. *Social Development*, *8*, 335–348.

Smetana, J. G., Kelly, M., & Twentyman, C. T. (1984). Abused, neglected, and nonmaltreated children's conceptions of moral and socio-conventional transgressions. *Child Development*, *55*, 277–287.

Smetana, J. G., Toth, S. L., Cicchetti, D., Bruce, J., Kane, P., & Daddis, C. (1999). Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. *Developmental Psychology*, *35*, 269–281.

Smith, D. W., Letourneau, E. J., Saunders, B. E., Kilpatrick, D. G., Resnick, H. S., & Best, C. L. (2000). Delay in disclosure of childhood rape: Results from a national survey. *Child Abuse & Neglect*, *24*, 273–287.

Sullivan, P. S., Dunham, R. G., & Alpert, G. P. (1987). Attitude structures of different ethnic and age groups concerning police. *Journal of Criminal Law and Criminology*, *78*, 177–196.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law & Human Behavior*, *26*, 395–415.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law & Human Behavior*, *28*, 411–435.

Taylor, T. J., Turner, K. B., Esbensen, F., & Winfree, L. T. (2001). "Coppin" an attitude: Attitudinal differences among juveniles toward police. *Journal of Criminal Justice*, *29*, 295–305.

Toth, S. L., Cicchetti, D., Macfie, J., Rogosch, F. A., & Maughan, A. (2000). Narrative representations of moral-affiliative and conflictual themes and behavioral problems in maltreated preschoolers. *Journal of Child Clinical Psychology*, *29*, 307–318.

Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science*, *3*, 92–109.

Wagland, P., & Bussey, K. (2005). Factors that facilitate and undermine children's beliefs about truth telling. *Law and Human Behavior*, *29*, 639–655.

Watson, A. J., & Valtin, R. (1997). Secrecy in middle childhood. *International Journal of Behavioral Development*, *21*, 431–452.

Woodcock, R., & Johnson, M. (1989). *Woodcock-Johnson Tests of Achievement (WJ-R)*. Allen, TX: DLM Teaching Resources.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1729–1738

Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons

Olga Kochukhova and Gustaf Gredebäck
Uppsala University

This study relies on eye tracking technology to investigate how humans perceive others' feeding actions. Results demonstrate that 6-month-olds ($n = 54$) anticipate that food is brought to the mouth when observing an adult feeding herself with a spoon. Still, they fail to anticipate self-propelled (SP) spoons that move toward the mouth and manual combing actions directed toward the head. Ten-month-olds ($n = 54$) and adults ($n = 32$) anticipate SP spoons; however, only adults anticipate combing actions. These results suggest that goal anticipation during observation of feeding actions develops earlier and is less dependent on directly perceived actions than goal anticipation during observation of other manual actions. These results are discussed in relation to experience and a possible phylogenetic influence on perception and understanding of feeding.

One of the earliest goal-directed actions that we experience in our life is feeding. During the first months of life, a regular intake of food is of the highest priority to a young infant. In addition to providing necessary nutrients, it is an arena for comfort and social support, all of which are important aspects of early development. Infants generally start breastfeeding or drinking formula from a bottle. However, this early monopoly on liquids is quickly broken as solids are introduced. In fact, Lepage, Moisan, and Gaudet (2004), reported that 99.8% of Canadian 6-month-olds and 60% of 4-month-olds had been fed other types of food than breast milk; as many as 93.5% of the 6-month-olds had been given cereals. In a similar study, Traveras et al. (2004) reported that only 24% of all 6-month-olds were fed exclusively with breast milk.

In addition to engaging in feeding, young infants (from 3 to 6 months) often place objects inside their mouths, spending on average 4–20 min per day performing such actions (Juberg, Alfano, Coughlin, & Thompson, 2001). Over these months, infants develop substantial mouth–hand coordination. Starting at 5 months of age, infants anticipate their hand approach through predictive mouth opening (Lew & Butterworth, 1997). However, it is only at 9 months that infants start picking up spoons and placing them correctly in their mouths (McCarty & Keen, 2005). Jointly, the importance of feeding combined with infants' extensive hand–mouth coordination ensure that feeding (or placing objects inside one's mouth) is a special (and perhaps unique) class of actions. Despite this, surprisingly little is known about how infants perceive and interpret observed feeding actions (for an exception, see Gredebäck & Melinder, 2010). In contrast, recently, several studies have been published that pursue the development of action understanding during observation of others' manual actions (Falck-Ytter, Gredebäck, & von Hofsten, 2006; Flanagan & Johansson, 2003; Gredebäck, Stasiewicz, Falck-Ytter, Rosander, & von Hofsten, 2009).

Numerous studies have relied on eye tracking and anticipatory gaze shifts to investigate the connection between action production and action understanding in both adults and infants. These studies demonstrate that adults look at the goal of their own manual actions before they are completed (Johansson, Westling, Bäckström, & Flanagan, 2001; Land, Mennie, & Rusted, 1999; Lee, Yound, Reddish, Lough, & Clayton, 1993). Infants and adults also anticipate the goal of others' manual actions (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003; Gredebäck et al., 2009). A recent illustration of this finding was reported by Falck-Ytter et al. (2006). In this study, 6- and

---

The study was supported by the Swedish Research Council (421-2006-1794).

Correspondence concerning this article should be addressed to Olga Kochukhova, Department of Psychology, Uppsala University, Blåsenhus, Von Kraemers allé 1, Uppsala 75142, Sweden. Electronic mail may be sent to olga.kochukhova@psyk.uu.se.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0009

12-month-old infants, as well as adults, were presented with an actor who either placed toys inside a bucket or sat passively as the toys moved on their own to the bucket (the agent's head was not visible). The results indicated that 1-year-olds and adults looked at the bucket, before the toys arrived, but only when a human model moved the toys. Six-month-old infants were unable to anticipate the goal when the model moved the toys. The authors note that infants become increasingly proficient at placing objects inside containers between 6 and 12 months of age (Bruner, 1970) and argue that this experience drives the development of anticipatory gaze shifts.

These results are consistent with habituation studies emphasizing that both experience and the perception of a human agent are important for the ability to encode the goal of manual actions (Sommerville, Woodward, & Needham, 2005; Woodward, 1998, 1999; Woodward & Guajardo, 2002). For example, Woodward (1998) presented 5-, 6-, and 9-month-old infants with a model who reached for one of two toys. After habituation, the toys changed place and the model reached either for the same toy in a new location or for a new toy at the old location. All participants looked longer when the model reached for the new toy but not when the model reached along a new path getting the old toy. This selective attention to goals was only present during observation of human reaching and not when infants observed similar actions being performed by a mechanical claw. One factor that influences this enhanced goal processing is infants' own experience in performing the same task. This was demonstrated by Sommerville et al. (2005), who obtained results similar to Woodward (1998) with 3-month-old infants, when they were trained on a reaching task prior to habituation.

At the same time, by using habituation procedures, investigators have demonstrated that young infants are sensitive to the rationality of movement with a wide range of self-propelled (SP) objects (Csibra, 2008; Csibra, Bíró, Koós, & Gergely, 2003; Kamewari, Kato, Kanda, Ishiguro, & Hiraki, 2005). For example, in a recent study by Csibra (2008), 6.5-month-old infants were habituated to movies of SP boxes that moved past an obstacle to reach a goal (functional movement). After this habituation phase, the obstacle was removed. In these test trials, infants looked longer if the agent continued on the old, nonfunctional path but not if the agent changed to a new functional path to reach the goal. The results of this study are interpreted in favor of the teleological stance (Gergely & Csibra, 2003), suggesting that infants distinguish functional and nonfunctional events from each other based on how objects move relative to the environment. According to teleological stance perspective, action understanding is mediated by visual properties of the observed events, including the path taken by the moving entity and its relation to the goal. In addition, infants are sensitive to the rationality of observed events (e.g., the degree to which objects follow straight or optimal paths to reach a goal) and expect objects to move, and actions to be carried out, in a rational manner. These assessments are based on visual properties of perceived events and are neither dependent on one's own experience (performing the same action) nor require the presence of a human actor (see Gergely & Csibra, 2003; Kiraly, Jovanovic, Prinz, Aschersleben, & Gergely, 2003).

The findings by Falck-Ytter et al. (2006) and Woodward (1998) demonstrate that infants' experiences are important for their ability to encode and anticipate the goals of observed events. However, infants may rely on visual properties of the observed motion and the relation between the moving object and the goal (Gergely & Csibra, 2003). Given the solid empirical findings that exist in support of both these accounts it is unlikely that only one of these mechanisms is present and influences infants' understanding of others', actions. However, little is known about their relative importance for guiding anticipatory gaze shifts and how their relative importance changes over time.

This study attempts to enhance our understanding of how infants perceive and interpret actions performed by others. In addition, we strive to map out the relative importance of two processes (experience and the more experience-independent rationality assessments suggested by the teleological stance). Are infants' goal-directed gaze shifts *primarily* guided by experience with similar actions or by visual motion properties and the relation between the moving entity and the goal?

To answer these questions, we developed two experiments. In the first experiment, we present infants with normal feeding actions (an action that infants are well acquainted with) and feeding events that do not include a manual action (SP spoons). The second experiment presents infants with a different manual action, which holds similar motion characteristics as the feeding action (presented in Experiment 1), but in this case, the action was largely unknown to the infants (e.g., a model combing her hair). Both experiments record infants' gaze as they observe these events. Infants' ability to anticipate the current action goal was compared over conditions.

Experiment 1

Six- and 12-month-old infants as well as adults were presented with one of two videos featuring feeding actions. In one condition, the feeding action corresponded with participants' own experience (human feeding [HF] condition), whereas the other condition featured a novel feeding event that included the same goal but lacked the manual components required to perform the feeding task; in this case, the spoon moved on its own toward the mouth (SP condition). The assumption that goal anticipation during observation of manual actions is *primarily* driven by experience with similar tasks suggests faster goal fixations (more anticipations) during normal feeding compared to matched SP events. The alternative assumption, proposed by teleological stance theory, where goal anticipations are primarily guided by visual properties of the motion and its relation to the goal suggests equal goal fixation latencies to both manual feeding actions and SP spoons, because these properties are matched across conditions.

We included 6-month-olds because at this age, infants already have some experience being fed, and, at the same time, they fail to anticipate other manual actions (Falck-Ytter et al., 2006). In comparison to 6-month-olds, 12-month-olds have ample experience being fed, they are able to place spoons correctly into their mouth (McCarty & Keen, 2005), and anticipate other manual actions (Falck-Ytter et al., 2006). The novelty of this task warranted the inclusion of an adult control group to frame infants' development.

Method

Participants. Thirty-six 6-month-old infants (24 females, mean age = 187 days, *SD* = 8), thirty-six 10-month-old infants (16 females, mean age = 303 days, *SD* = 9), and 24 adults (13 females, mean age = 31.4 years, *SD* = 8) participated in Experiment 1. Eighteen infants were included in each experimental condition. All infants were from middle-class Caucasian families. Only 1 participant was excluded, a 10-month-old infant, because of fussiness. All infants were full-term and healthy. Mean experience with spoon-feeding in 6-month-old group was 1 month 7 days (*SD* = 14 days). Only three 6-month-old infants had no experience being fed with a spoon. All 10-month-olds had experience being fed with a spoon; mean experience with spoon-feeding equaled 5 months 12 days (*SD* = 21 days) and did not differ between conditions. Three 10-month-old infants in the SP condition and two 10-month-olds in the manual condition could successfully use a spoon to feed themselves.

Stimuli and procedure. All infants were recruited by mail, whereas adults were recruited from the university campus. All signed a consent form prior to participation and received a gift voucher (€14). The study was approved by the Regional Ethics Committee according to the 1964 Declaration of Helsinki.

Prior to participation, we asked parents several questions about their infant: (a) Do you feed your child with a spoon? (b) When did you start to feed your child with a spoon? and (c) Can your child put a spoon with food into her or his mouth (feed her- or himself successfully)? After these questions, infants were seated in a safety car seat on the parent's lap. Their gaze was measured with a Tobii 1750 (Stockholm, Sweden) near infrared gaze tracker system (subjects' eyes approximately 60 cm from the monitor). The camera records near infrared reflections of both eyes at 50 Hz (accuracy = 0.5°, spatial resolution = 0.25°) while the stimuli is presented on a 17-in. screen.

During the 9-point calibration phase, infants watched a blue-and-white sphere over a white-dotted black background (Gredebäck & von Hofsten, 2004) expand and contract (extended diameter = 3.3°) in synchrony with a sound. In cases where infants were inattentive, a small frame with a cartoon video was shown until infants attended again to the monitor. Following this calibration, participants were presented with a context movie (identical in both conditions) and one of two possible sets of experimental movies (each participant only viewed one set of experimental movies; i.e., they participated in only one condition).

During the context movie, a female peeled and cut a banana and placed its pieces on a plate (Figure 1a). Duration of the context movie was 29 s. This movie was followed by five identical condition specific movies. In the *HF condition*, these movies started with the actor looking into the camera for 1 s with her hands on her lap. The actor then turned her attention to the plate by turning her head and eyes to the plate while simultaneously lifting her arm to take the spoon from the table. After the spoon was picked up, the actor took a piece of banana from the plate, and brought it to

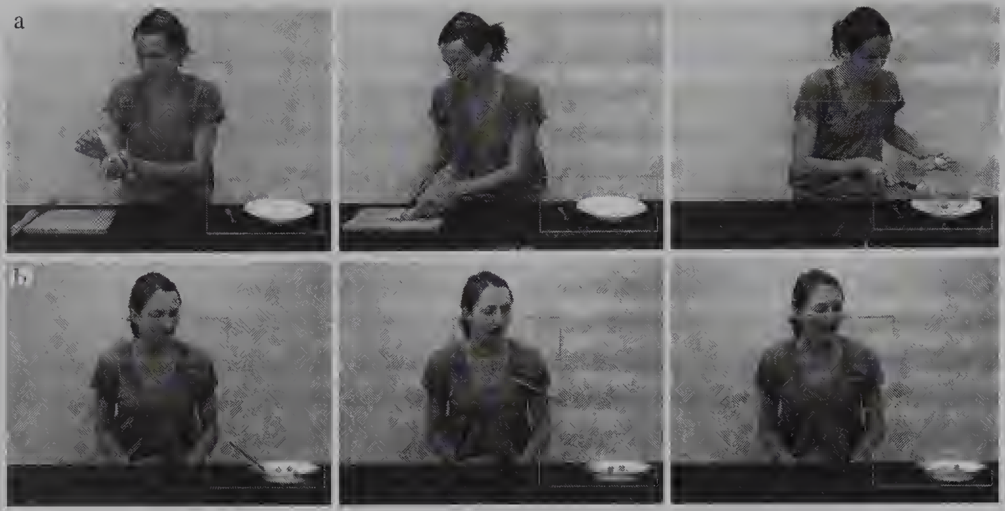


Figure 1. Snapshots from initial context movie (a) and the self-propelled (SP) condition (b). The upper rectangle (mouth area of interest [AOI]) covers the model's mouth, taking into account head movement during the movie; the lower rectangle depicts the location of the plate AOI.

her mouth. In the *SP condition*, the same sequence was repeated with the only exception being that the actor did not produce any manual actions; instead, the spoon flew by itself to the plate, took a piece of banana, and moved toward the actor's mouth (Figure 1b). In both conditions, the feeding action ended once the spoon had deposited the banana inside the actor's mouth and once the food had been eaten. Duration of each condition specific movie was 39 s; each movie included four feeding events (actions). Thus, each participant saw 20 feeding events during the experiment. The duration of the spoon's movement, trajectory of the spoon, the time when the model's mouth opened, and the location of the model's head were matched over conditions. All movies extended 22.5° × 19.5° of visual angle, the agent's head covered 4° × 5°, and the plate 4.3° × 1.7°. The angular distance between the plate and the mouth extended 14° and the vertical extent of the agent in the picture was 13°. The entire experimental session lasted no more than 5 min.

Data analysis. All analyses were performed using custom-made analysis programs in the Matlab (Mathworks Inc., Natick, MA) environment. Two areas of interest (AOIs) were defined (gray rectangles in Figure 1); one covered the starting position of the spoon (plate AOI) and one covered the area around the mouth of the actor (mouth AOI). Gaze was analyzed only during each spoon movement from the plate AOI to the mouth AOI. All AOIs had the same size and location; the mouth AOI covered the location of the mouth during all head movements performed by the actor. Participants' gaze shifts were included in the analysis only if subjects fixated the plate AOI for 200 ms (Gredebäck, Örnkloo, & von Hofsten, 2006), then shifted their gaze to the mouth AOI and fixated it for 200 ms, before the spoon left this area. Thus, participants could pay attention to the video sequence on a given trial and still be excluded from analysis, if attention was not directed to the feeding action (in accordance with the criteria mentioned before).

The timing of the infants' gaze shifts to the mouth AOI was compared to the time when the actor started to open her mouth after the spoon left the plate AOI. Gaze shifts above 0 were considered predictive, indicating that participants shifted their gaze to the mouth AOI before the actor started to open her mouth.

Results were analyzed through factorial analyses of variance (ANOVAs) with age (6-, 10-month-olds, and adults) and condition (HF and SP) as independent variables. Significant effects were further explored using Tukey honestly significant difference (HSD) post hoc tests. The degree to which the latency of participants gaze shifts differed significantly from 0 (separating predictive and reactive conditions) was estimated with Bonferroni corrected single-sample *t* tests.

Results

Six-month-old infants attended to feeding actions in 40% (*SE* = 5) of trials in the SP condition and in 42.5% (*SE* = 5) of trials in the HF condition. Mean amount of anticipations, made during these trials, was 37% (*SE* = 6) in the SP condition and 49% (*SE* = 6) in the HF condition. Ten-month-old infants and adults followed the feeding action in 52% (*SE* = 5) and 64% (*SE* = 6) of trials in the SP condition versus 46% (*SE* = 5) and 76% (*SE* = 6) of trials in the HF condition, respectively. Ten-month-olds evidenced anticipations on 45% (*SE* = 7) of attended trials in the SP condition and on 62% (*SE* = 6) of attended trials in the HF condition. For adults, these numbers equaled 62% (*SE* = 6) and 71% (*SE* = 6), respectively.

Preliminary tests did not indicate any order effect in either of the conditions ($p > .05$); data were therefore aggregated across presentation order. A factorial ANOVA with number of attended trials as the dependent variable demonstrated a significant main effect of age, $F(2, 90) = 14$, $p = .001$; post hoc tests differentiated the higher attention rates to the action in adults from attention rates in both 6- and 10-month-old infants ($p = .001$).

More important, the latency of participants' gaze shifts to the mouth during feeding actions differed depending on the age of the participant, $F(2, 90) = 6.3$, $p = .005$. Post hoc tests demonstrated that 6-month-old infants shifted their gaze to the model's mouth much later than 10-month-olds ($p = .006$) and adults ($p = .001$). Latencies also differed between the two conditions, $F(1, 90) = 7.5$, $p = .01$, with faster latencies during HF ($p = .005$). Single-sample *t* tests demonstrate that 6-month-olds predicted the goal of manual feeding actions, $t(17) = 3.6$, $p = .002$, but not SP feeding (*ns*). Ten-month-olds and adults anticipated the goal of feeding action during both manual, 10-month-olds: $t(17) = 4.2$, $p = .001$; adults: $t(11) = 4.8$, $p = .001$; and SP feeding, 10-month-olds: $t(17) = 3.8$, $p = .01$; adults: $t(11) = 3.6$, $p = .005$ (see Figure 2).

Discussion

Experiment 1 demonstrates that 6-month-old infants anticipate the goal of normal feeding actions and track SP spoons in a reactive manner. At the same time, all age groups showed faster goal fixation latencies to normal feeding than during observation of SP spoons. Two lines of evidence suggest that these differences can be related to the observers' own experience with similar tasks.

First, the fact that 6-month-olds anticipate normal feeding actions (but not SP spoons) supports the predictions derived from the experience account. According to this perspective, experience facilitates action understanding and goal anticipation of tasks that closely accord with the infants own experience. The fact that both 10-month-olds and adults demonstrated faster anticipations in the manual feeding condition compared to the SP condition suggests that the same general mechanism holds an influence on action understanding and goal anticipation throughout development. At the same time, this finding does not support the assumption proposed by the teleological stance theory that young infants rely predominantly on visual properties of the motion and its relation to the goal when observing ongoing manual actions.

However, there are differences between the manual and SP conditions (see also Falck-Ytter et al., 2006) that do not relate solely to the presence or absence of human actions and the familiarity of the task. In the manual feeding condition both the arm

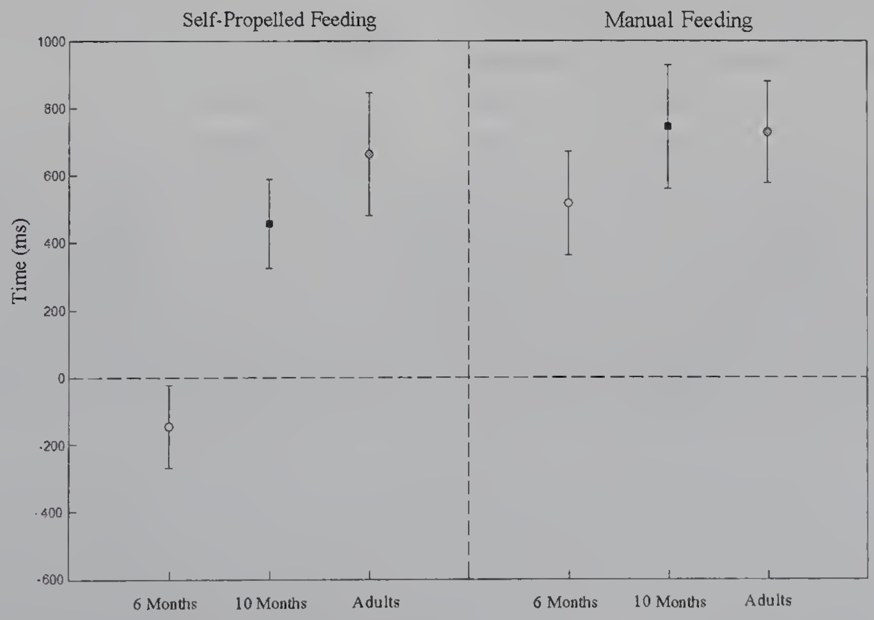


Figure 2. Gaze performance during observation of actions in both experimental conditions.
Note. Vertical axis depicts timing of gaze arrival at the mouth area of interest (AOI) relative to the time of actor's mouth opening during the action. The start of mouth opening is represented by the horizontal line at 0 ms. Positive values correspond to early arrival of gaze at the target area. Error bars represent standard errors of the mean.

and the spoon move, creating prominent motion cues in the direction of the goal, whereas the SP spoon provides the observer with a similar, but smaller, degree of motion. It is possible that this difference in visible motion is a mediating factor that influences the latency of goal-directed gaze shifts, with more predictions in response to large moving objects (HF condition) than during observation of a small moving object (SP condition). At the same time, it should be noted that arm movements alone are not sufficient to produce anticipatory gaze shifts. In a recent article, Gredebäck et al. (2009) demonstrate that infants only anticipate the goal of functional manual actions directed toward salient goals. Simply observing a moving hand, without a clear connection to a goal, did not result in anticipatory gaze shifts. Based on this finding, it is unlikely that the saliency of the movement is the sole cause of the differences reported in Experiment 1. However, the possibility still exists that motion becomes a mediating factor that adjusts the latency of goal-directed gaze shifts in the presence of a salient goal. We refer to this alternative account as the *degree of motion hypothesis*.

Second, the suggestion that action understanding (as measured by goal anticipations) is influenced by experience receives additional support through comparing the current findings to those obtained by Falck-Ytter et al. (2006). According to the literature reviewed in the Introduction (and the parental reports described in the Results section of Experiment 1), infants possess ample experience participating in manual feeding actions before 6 months of age. Manual containment actions (Falck-Ytter et al., 2006), however, are not readily performed by infants before they are 7–9 months of age (Bruner, 1970). Accordingly, 6-month-olds anticipate manual feeding actions but not containment actions. Goal anticipation of the latter action only occurs once infants have become proficient at placing objects inside containers at 12 months of age. It is tempting to attribute these differences in onset of anticipatory gaze shifts to infants' own experience with the specific task being observed. However, it is possible that goal saliency represents a mediating factor that differentiates the onset of anticipatory gaze shifts (between the two studies) irrespective of experience. According to this alternative account (referred to as the *saliency hypothesis*), young infants anticipate the goal of feeding actions from an early age because the human head is a highly salient goal whereas goal anticipation of manual containment actions emerges later because the goal of this action is less salient (a bucket).

Experiment 2 is designed to test both of these alternative hypotheses, by presenting 6- and 10-month-old infants and adults with a goal-directed manual action that only adults have experience performing, namely, combing. Manual combing actions provide the observer with motion information from the hand and the comb (serving as a control for the relative motion hypothesis). At the same time, combing actions are directed toward the head (providing a control for the saliency hypothesis). Both of these alternative hypotheses suggest that infants will be able to anticipate the goal of this manual action. However, if infants' own experience is the primary cause of action understanding and goal anticipation, we predict that only the adult group will be able to anticipate the goal of this action. This prediction is based on parental reports, indicating that infants' experiences with such tasks are limited (see the Participants section of Experiment 2).

Experiment 2

Experiment 2 presents 6- and 10-month-old infants, as well as adults, with movies of an adult human who picks up a comb and combs her hair. The main purpose of this experiment is to test the validity of the claim that experience mediates a strong influence on infants' action understanding and goal anticipations. To allow comparisons with Experiment 1, both the sequence of actions and the timing of each action were matched to the conditions presented in Experiment 1.

Method

Participants. Eighteen 6-month-olds (eight females, mean age = 184 days, *SD* = 5), an equal number of 10-month-olds (12 females; mean age = 304 days, *SD* = 9), and 8 adults (5 females; mean age = 29 years, *SD* = 7) participated in Experiment 2. Three additional 6-month-olds and 1 additional 10-month-old infant were excluded because of fussiness. All infants were full-term and healthy; the majority of families came from a Caucasian middle-class background. Parents of two 6-month-old infants and four 10-month-old infants reported that their children on occasion saw them comb their hair. None of the participating infants had experience combing their own hair.

Stimuli and procedure. The procedure in this experiment was the same as in Experiment 1 with one exception. Participants' gaze was measured with Tobii T120 near infrared gaze tracker system that records near infrared reflections of both eyes at 60 or 120 Hz sampling frequency as participants watched a 17-in. monitor (accuracy = 0.5°, spatial resolution < 0.3°). In this experiment, gaze was sampled at 60 Hz; this is a slightly higher sampling rate than in Experiment 1 (50 Hz).

Prior to participation, the following questions were asked of parents: (a) Did your child ever see anybody comb their hair? If yes, how often? (b) Do you comb your child's hair? How often? and (c) Can your child comb his or her own hair? If yes, how often? Following these questions, participants went through the same procedure that is outlined in Experiment 1.

The context video demonstrated the same actor as in Experiment 1 (see Figure 3). In this first video, the model opens a small bag placed on the table, takes out a comb, and places it next to a plate. The model thereafter takes a cup of water and pours the water into the plate. The second movie starts with the model looking into the camera for 1 s with hands on her lap. The model then turns her head and eyes to the plate simultaneously raising her arm to grab the comb. After the comb is picked up the actor dips the front end of the comb in the water while holding the comb by its handle. Following these preparations (similar to the preparations performed in Experiment 1), the model brings the comb to her hair and makes three vertical combing movements after which the comb is brought to the plate again and the same action sequence is repeated three more times. Both spatial and temporal aspects of the stimuli are matched to the stimuli presented in Experiment 1.

Data analysis. All analyses in this experiment were performed using the same analysis tools (with the same criteria) used to analyze the results of Experiment 1. The latencies of participants' goal-directed gaze shifts were related to the movement of the arm and comb toward the model's head. This was done in two separate analyses. In one analysis, the latency of participants' first fixation anywhere on the head was related to the time when the comb first entered the area covering the head (head AOI, see Figure 3a). In the second analysis, the latency of participants' first fixation on the part of the hair that was combed was related to the time when the comb first entered this area (hair AOI; see Figure 3b).

Results

Six-month-old infants attended to the combing action in 38% (*SE* = 5) of all trials. Mean amount of anticipation for these trials equaled 24% (*SE* = 6). Ten-month-old infants and adults attended to 45% (*SE* = 7) and 96% (*SE* = .8) of trials, respectively. Amount of anticipation for these trials equaled 36% (*SE* = 6) for 10-month-olds and 78% (*SE* = 3) for adults.

We could not observe any effect of presentation order across the trials ($p > .05$), so the data were compiled across order. ANOVA analysis performed on the goal fixation latencies produced to the head AOI demonstrated significant age differences, $F(2, 41) = 6.5$, $p = .005$. Post hoc tests demonstrated that adults fixated the goal faster than both infant groups ($p = .01$). According to single-sample t tests, adults were the only group who produced predictive gaze shifts to the head AOI, $t(7) = 3.7$, $p = .008$.

Similar analyses performed on the gaze shifts produced to the hair AOI revealed the same pattern. Again ANOVA analysis demonstrated

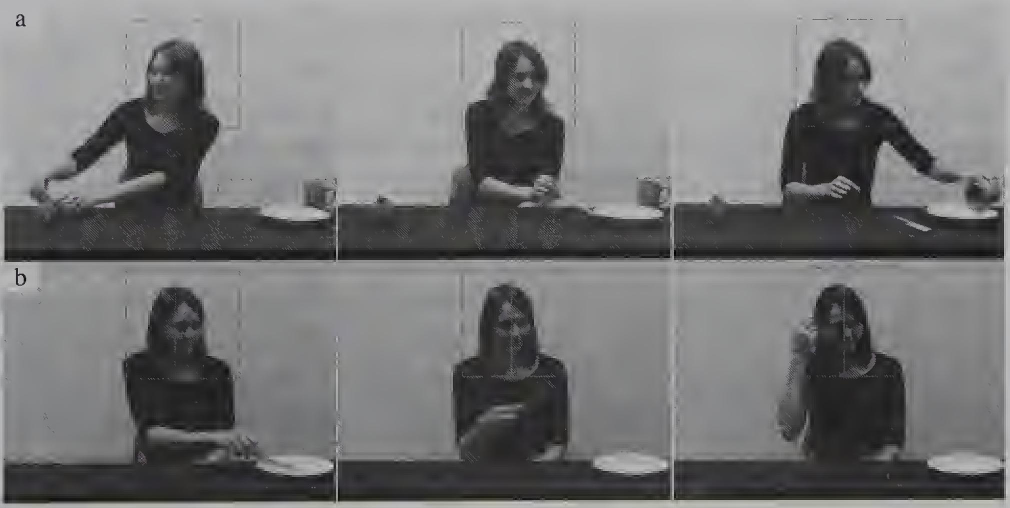


Figure 3. Snapshots from context movie (a) and the combing action (b).
Note. The upper rectangles (dotted line) represent placement of head area of interest (AOI). Narrow upper rectangle (solid line at b) demonstrates placement of hair AOI. The lower rectangle (dotted line) depicts the location of the plate AOI.

significant age differences, *F*(2, 41) = 18, *p* = .001 (see Figure 4); post hoc tests revealed much faster gaze fixations demonstrated by adults (*p* = .001). Moreover, adults were again the only group who produced predictive gaze shifts to the goal area, hair AOI, *t*(7) = 2.7, *p* = .03.

Discussion

Experiment 2 presented participants with a normal, goal-directed, and functional action that maintained the spatial and temporal properties of the feeding action presented in Experiment 1. Despite this, only adults were able to anticipate the goal of this action. This late onset of anticipatory gaze shifts in Experiment 2 does not support the degree of motion hypothesis. The observed action includes a large degree of motion from the arm and comb (matched to Experiment 1); however, this information is not sufficient to allow infants to anticipate the current action goal. Nor does it support the saliency hypothesis, according to which infants might produce an early onset of anticipatory gaze shifts when observing actions directed toward a human head and a delayed onset when observing action directed toward less salient stimuli. In this respect, it should be emphasized that the current finding does not falsify the general assumption that goal saliency is an important aspect of action understanding and goal anticipation. It does, however, suggest that differences in goal saliency cannot account for the large variation in onset of anticipatory gaze shifts observed during observation of feeding (Experiment 1), containment actions (Falck-Ytter et al., 2006), and combing (Experiment 2).

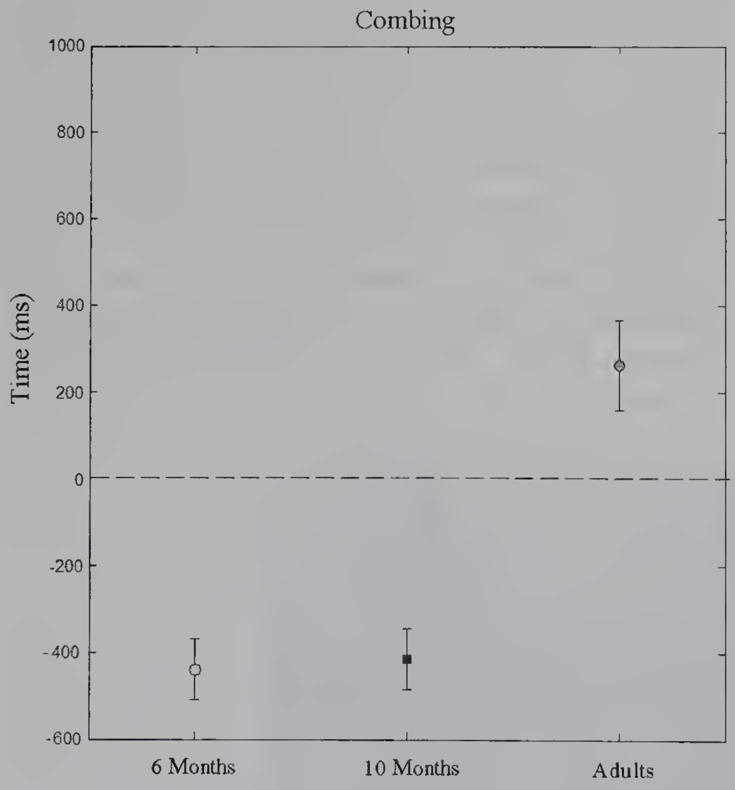


Figure 4. Gaze performance during observation of combing action.
Note. Vertical axis depicts timing of gaze arrival at the hair area of interest (AOI) relative to the time when the comb first entered the Hair AOI. The time when the comb first came into the hair AOI is represented by the horizontal line at 0 ms. Positive values correspond to early arrival of gaze at the target area. Error bars represent standard errors of the mean.

The two different analyses used in Experiment 2 serve somewhat different purposes. By comparing latencies of goal fixations with respect to the part of the hair that was actually combed, we were able to make direct comparisons to Experiment 1. In both experiments, these AOIs cover only the essential goal area (where the spoon and comb interact with the human goal) while maintaining an identical AOI size over the two experiments. It is, however, possible to question the appropriateness of this matched AOI for analyzing the data of Experiment 2. First, infants might find it difficult to anticipate exactly which area of the head the model intends to comb. As such, this analysis might miss a number of anticipations directed toward other parts of the hair. In a similar manner, some infants might believe that this action was directed toward the mouth or some other part of the head, simply because they may have misunderstood the purpose of the action. To control for this possibility, we also reported the latency of gaze shifts to the entire head. As can be seen in Experiment 2, this possibility did not systematically influence the results because infants made few anticipatory gaze shifts in this condition.

General Discussion

This is the first study to demonstrate goal anticipation by 6-month-old infants during observation of manual actions. In addition, 6-, 10-month-olds, and adults all demonstrate faster gaze shifts to the goal of manual actions than during observation of matched SP events. At the same time, both 6- and 10-month-olds fail to anticipate a highly similar combing action with which they have no (or at least limited) experience. Adults, who have much experience with similar events, anticipated the goal of the later action. All of these findings support the experience account, suggesting that our own experience plays an important role when we perceive and anticipate the goal of other people's actions.

The tight connection between the onset of anticipatory gaze shifts and experience performing the same manual tasks in infants suggests that observational experience is not sufficient to guide anticipatory gaze shifts. Instead, it seems that infants rely on a direct matching process that simulates action goals by relating observed events to corresponding manual action plans (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003). Neither of these findings supports the assumption that anticipatory gaze shifts are guided by well-formedness criteria, as defined within the teleological stance framework (Gergely & Csibra, 2003).

Contrary to prior findings, both 10-month-olds and adults anticipated the goal of SP feeding actions (however, with longer latencies than during observation of manual actions). Previous studies have only demonstrated anticipatory gaze shifts to functional goal-directed manual actions, not to SP events (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003) or to nonfunctional arm movements (Gredebäck et al., 2009). These findings suggest that the link between manual feeding actions and anticipation during observation of feeding is less dependent on actually perceiving the manual action. As such, it could be argued that the goal anticipation during observation of feeding is more flexible, or less dependent on, visual information about the hand and arm than what has been demonstrated for other manual actions (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003; Gredebäck et al., 2009).

Of course, it could be claimed that 10-month-olds and adults apply teleological reasoning to anticipate SP-feeding events. However, this suggestion would not be sufficient to explain the faster goal fixation latencies observed during manual feeding actions. Moreover, it would not be able to account for the synchrony in onset between the ability to anticipate the goal of others' actions (Falck-Ytter et al., 2006) and infants' own manual abilities (such as placing objects inside their mouth; Experiment 1). At the same time, it is important to emphasize that there is no empirical evidence (or theoretical claims) that would support the notion that well-formedness criteria (and teleological reasoning) should be especially directed toward anticipating *feeding* actions. Such task specificity is not consistent with the broad applicability and flexibility of teleological reasoning as described by Csibra and Gergely (2007).

Instead, we suggest two alternative, and possibly complementary, explanations for our findings. First, feeding is a biologically rooted action that might hold special meaning to both humans and other species. In other words, understanding of feeding might be phylogenetic primed, enhancing an observer's ability to anticipate feeding actions even when the entire manual action is not clearly observable (SP condition of Experiment 1). Indeed, from an evolutionary point of view it would be of the highest priority to understand the intentions of others in situations that are directly connected to one's own physical survival, such as feeding. It is frequently reported that primates in the wild pay attention to what others are eating (Perry & Manson, 2003) and where they are feeding (Kazahari & Agetsuma, 2008). For example, mantled howler monkeys must carefully watch what their mothers eat in order not to be poisoned (Perry & Manson, 2003). Observation of others is also important when usual food depletes. This has been demonstrated at several Japanese macaque sites where adult males' discovery of new food leads to transmission of the habit to other macaques' (Perry & Manson, 2003). Based on these findings, we assume that humans might be able to extend the range of feeding events that can be mapped on to our own motor capabilities, through processes deeply rooted in our phylogenetic history. The possible evolutionary advantage associated with such an enriched action understanding of feeding is easily perceived: It would allow better access to food and a competitive edge that might enhance chances of survival when resources are scarce.

A second reason might be infants' high frequency of experience being fed. Infants are exposed to feeding, place objects inside their mouths, and develop their own hand–mouth coordination from a very early age. The high frequency of such experience at an early age might make it easier for infants to connect observed events (manual actions and moving spoons) with their own enactment of the goal (opening the mouth to consume the food) and their own experience with putting things into the mouth. All of these experiences might promote an understanding of feeding that exceeds even adults' understanding of other manual actions. Future studies are necessary to investigate how each of previously named factors influences our understanding of others' actions and especially our ability to anticipate the goal of manual feeding actions.

At the same time, if we reason more broadly, it is clear that adults are able to use more than one proactive strategy, when interacting with the world. The use of experience-only strategy would exclude the possibility of making correct predictions in completely new situations. However, taking into account the results of this study, it seems that during development, relying on one's own experience is a leading strategy for perceiving and anticipating the goal of other people's actions.

References

Bruner, J. S. (1970). The growth and structure of skill. In K. Connolly (Ed.), *Mechanisms of motor skill development* (pp. 63–94). New York: Academic Press.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*, 705–717.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of action productivity. *Cognitive Science*, *27*, 111–133.

Csibra, G., & Gergely, G. (2007). ''Obsessed with goals'': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*, 60–78.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*, 878–879.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *14*, 769–771.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*, 287–292.

Gredebäck, G., & Melinder, A. Infants understanding of everyday social interactions: A dual process account. *Cognition*, *114*, 197–206.

Gredebäck, G., Örnkloo, H., & von Hofsten, C. (2006). The development of reactive saccade latencies. *Experimental Brain Research*, *173*, 159–164.

Gredebäck, G., Stasiewicz, D., Falck-Ytter, T., Rosander, K., & von Hofsten, C. (2009). Hand configuration and goal type modulate goal directed gaze shifts in 14-month-old infants. *Developmental Psychology*, *45*, 1190–1194.

Gredebäck, G., & von Hofsten, C. (2004). Infants' evolving representation of moving objects between 6 and 12 months of age. *Infancy*, *6*, 165–184.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, *21*, 6917–6932.

Juberg, D. R., Alfano, K., Coughlin, R. J., & Thompson, K. M. (2001). An observational study of object mouthing behavior by young children. *Pediatrics*, *107*, 135–142.

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development*, *20*, 303–320.

Kazahari, N., & Agetsuma, N. (2008). Social factors enhancing foraging success of a wild group of Japanese macaques (Macaca fuscata) in a patchy food environment. *Behaviour*, *145*, 843–860.

Kiraly, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition*, *12*, 752–769.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328.

Lee, D. N., Yound, D. S., Reddish, P. E., Lough, S., & Clayton, T. M. (1993). Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology A*, *35*, 333–336.

Lepage, M.-C., Moisan, J., & Gaudet, M. (2004). What do Québec children eat during their first six months? *Canadian Journal of Dietetic Practice and Research*, *63*, 106–113.

Lew, A. R., & Butterworth, G. (1997). The development of hand-mouth coordination in 2- to 5-month-old infants: Similarities with reaching and grasping. *Infant Behavior and Development*, *20*, 59–69.

McCarty, M. E., & Keen, R. (2005). Facilitating problem-solving performance among 9- and 12-month-old infants. *Journal of Cognition and Development*, *6*, 209–228.

Perry, S., & Manson, J. H. (2003). Traditions in monkeys. *Evolutional Anthropology*, *12*, 71–81.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*, B1–B11.

Traveras, E. M., Scanlon, K. S., Birch, L., Rifas-Shiman, S. L., Rich-Edwards, J. W., & Gillman, M. W. (2004). Association of breastfeeding with maternal control of infant feeding at age 1 year. *Pediatrics*, *114*, 577–583.

Woodward, A. L. (1998). Infants selective encode the goal of an actor's reach. *Cognition*, *69*, 1–34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, *22*, 145–160.

Woodward, A. L., & Guajardo, J. J. (2002). Infant's understanding of the point gesture as an object-directed action. *Cognitive Development*, *17*, 1061–1084.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1739–1752

Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning

Annie Bernier
University of Montreal

Stephanie M. Carlson
University of Minnesota

Stéphanie Bordeleau and Julie Carrier
University of Montreal

The aim of this report was to investigate the prospective links between infant sleep regulation and subsequent executive functioning (EF). The authors assessed sleep regulation through a parent sleep diary when children were 12 and 18 months old ($N = 60$). Child EF was assessed at 18 and 26 months of age. Higher proportions of total sleep occurring at night time, at both 12 and 18 months, were related to better performance on executive tasks, especially those involving a strong impulse control component. Most relations held above family socioeconomic status, prior mental development and concurrent verbal ability. These findings add to previous results with school-age children in suggesting that sleep favors the development of higher order cognitive functions requiring prefrontal cortex involvement.

A crucial developmental task in early childhood is the development of executive functioning (EF). EF refers to higher order cognitive processes, including impulse control, set shifting, and working memory, which aid in the monitoring and control of thought and action (Garon, Bryson, & Smith, 2008; Zelazo, Carter, Reznick, & Frye, 1997). The first signs of EF emerge in infancy as behaviors move from being monitored externally by caregivers to being regulated internally by the child. Individual differences in EF have often been linked to children's theory-of-mind understanding (for a review, see Zelazo, Carlson, & Kesek, 2008) as well as emotional and social competence, moral conduct, and school readiness and achievement (e.g., Blair & Razza, 2007; Carlson & Wang, 2007; Espy et al., 2004; Kochanska, Murray, & Harlan, 2000; van der Sluis, de Jong, & van der Leij, 2007). In contrast to this compelling set of findings, very little is known about the mechanisms that underlie the development of EF. Owing to the central role of frontal brain structures in EF (see Zelazo et al., 2008), and the close connections between early sleep regulation and brain development (Jones, 2005), this report examines the prospective relations between sleep regulation in infancy and subsequent EF.

Sleep is sometimes considered to be the single most important activity of the brain during early childhood (Dahl, 1996). In fact, infants spend over half of their first 2 years of life sleeping, with an average daily sleep duration of about 16–17 hr a day during the 1st month, gradually declining to 14–15 hr at approximately 6 months, and to 13 hr at 2 years of age (Coons & Guilleminault, 1992; Davis, Parker, & Montgomery, 2004; Iglowstein, Jenni, Molinari, & Largo, 2003). Hence, by 2 years of age, the average child has spent nearly 14 months asleep and only 10 months in waking activities (Dahl, 1996). During the same period, the brain undergoes a growth spurt, reaching 90% of its adult size, while myelination and synaptic pruning dramatically alter the landscape of neural networks (Nelson, Thomas, & de Haan, 2006). As

---

The research described in this article was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Fonds de Recherche en Santé du Québec, and the Fonds Québécois de Recherche sur la Société et la Culture to the first author. We gratefully acknowledge Natasha Whipple, Émilie Rochette, Natasha Ballen, Isabelle Demers, Jessica Laranjo, Célia Matte-Gagné, Chantal Mongeau, Marie-Pier Nadeau-Noël, and Nadine Marzougui for help with data collection. Special thanks go to the participating families of the *Grandir Ensemble* project who generously opened their homes to us.

Correspondence concerning this article should be addressed to Annie Bernier, Department of Psychology, University of Montreal, P.O. Box 6128, Downtown Station, Montreal, QC H3C 3J7, Canada. Electronic mail may be sent to annie.bernier@umontreal.ca.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0010

highlighted by Dahl (1996), the fact that evolution has favored the brain to spend considerably more time in a state of sleep than in all waking activities combined at such a time of rapid neural development suggests that sleep may serve a crucial function in brain development.

Consistent with this proposal, numerous studies have demonstrated that different indicators of sleep are associated with infants' and children's cognitive functioning (e.g., Becker & Thoman, 1981; Buckhalt, El-Sheikh, & Keller, 2007; Dearing, McCartney, Marshall, & Warner, 2001; Gomez, Bootzin, & Nadel, 2006; Scher, 2005; Steenari et al., 2003; Touchette et al., 2007). Notably, research has often found EF deficits among children suffering from sleep-disordered breathing (e.g., obstructive sleep apnea; see Beebe, 2006). These executive deficits are considered to be attributable to disrupted sleep, to hypoxemia, or to both (Beebe & Gozal, 2002). These findings with clinical populations appear to extend to the general population as well. Hence, although sleep deprivation in adults has a negative impact on performance on many psychomotor, attentional, and cognitive tasks (Boonstra, Stins, Daffertshofer, & Beek, 2007), it has often been observed that executive tasks are particularly affected by sleep deprivation (for reviews, see Dahl, 1996; Harrison & Horne, 1998; Jones & Harrison, 2001). Such findings have been interpreted as evidence for the notion that the prefrontal cortex, largely implicated in executive functions, may be particularly sensitive to the homeostatic processes underlying sleep regulation (e.g., Horne, 1993; Jones & Harrison, 2001).

Electrophysiological studies provide support for this interpretation by suggesting that sleep may have a localized homeostatic function of particular importance for frontal brain areas. The homeostatic process is a regulating mechanism whereby sleep pressure accumulates with time awake and dissipates during a sleep episode. In humans, the intensity and dynamics of slow-wave sleep (SWS) and slow-wave activity (SWA; spectral power between 0.5 and 4.0 Hz during non-rapid-eye-movement sleep) reflect the time course of the homeostatic process: More time awake produces higher levels of SWS and SWA, whereas more time asleep is associated with lower levels (Achermann, Dijk, Brunner, & Borbély, 1993). Of particular interest for our purposes are studies showing brain regional differences in this process. SWA is more prominent in frontal-central brain areas compared to posterior regions (Werth, Achermann, & Borbély, 1996), and the increase in SWS and SWA after sleep deprivation is larger in frontal than occipital regions (Cajochen, Foy, & Dijk, 1999). In addition, the impact of sleep deprivation on waking electroencephalogram is more prominent in frontal than posterior brain areas (Cajochen, Knoblauch, Krauchi, Renz, & Wirz-Justice, 2001). Thus, as argued by many (Beebe & Gozal, 2002; Dahl, 1996; Horne, 1993), frontal brain regions seem to have a higher need for recovery during sleep than more posterior areas.

Although little research has been conducted with nonclinical samples of children thus far, emerging evidence has begun to suggest there may be a specific link between EF and sleep in childhood. Sadeh, Gruber, and Raviv (2002) found that sleep fragmentation was negatively related to school-age children's performance on tasks requiring executive control, but not to simpler tasks of memory or reaction time. In a subsequent experimental study, Sadeh, Gruber, and Raviv (2003) found that three consecutive nights of at least 30 min of augmented sleep improved school-age children's performance on certain neurobehavioral tasks, including a "continuous performance task" used to assess attention and inhibition in preschoolers (Carlson, 2005). In another experimental study, Randazzo, Muehlbach, Schweitzer, and Walsh (1998) gave 8 preadolescents (10–14 years) one night of sleep restriction (5 hr in bed), and 8 others one night of sleep augmentation (11 hr in bed). They found that while none of the measures involving a low cognitive load (e.g., semantic memory) distinguished the two groups, children in the sleep augmentation group outperformed their sleep-restricted counterparts on the Wisconsin Card Sorting Test, often used as a measure of EF (Zelazo, Müller, Frye, & Marcovitch, 2003).

Hence, there is emerging evidence that sleep may play a specific role in EF in children. Thus far, however, these effects have been observed in children only 7 years and older. Given that major advances in EF take place between 1 and 6 years of age (Zelazo et al., 2008), and that neural density of the frontal lobes begins to decline at about 7 years of age (Huttenlocher, 2002), it appears that the potential impact of sleep on frontal/executive functions may be especially potent earlier in development, when the brain shows substantial dynamic plasticity in accord with experience (Singer, 1995). It is also noteworthy that all studies assessing child sleep and EF have done so concurrently, leaving the possibility open that the links are because of child sleepiness interfering with immediate performance on the more complex tasks. Longitudinal studies are important in teasing apart transient fatigue from persistent effects of sleep difficulties on

EF, which could be attributable to a direct impact of sleep on frontal brain development itself.

Finally, existing studies have used large cognitive batteries and numerous sleep indicators and found links between certain sleep parameters and some, but not all, expected cognitive indices. As highlighted by Buckhalt et al. (2007), drawing valid conclusions on presumed specific links between sleep and EF is complicated by the great variety of cognitive tasks used in different studies, and better construct specificity and measurement are needed. We further argue that appropriate measurement of EF with children should entail a developmental approach. Existing studies of sleep and EF have generally used measures designed to assess executive deficits in brain-injured adult patients. The developmental course and significance of such measures are unclear. In the last few years, developmental researchers have established age-appropriate measures of EF with identified developmental trends (e.g., Carlson, 2005; Hughes & Ensor, 2005; Zelazo et al., 2003), which are likely to be especially useful in drawing developmentally valid conclusions on the impact of sleep on EF in children.

In contrast, there is currently no consensus on the best indicators of sleep quality in infancy and childhood (Hayes, 2002; Jenni & Carskadon, 2007). In fact, a striking aspect of the sleep research involving infants and children is how different sleep parameters are associated with positive outcomes in different studies, and how the same sleep parameters show different (e.g., Anders, Keener, & Kraemer, 1985; Gertner et al., 2002) and sometimes opposite relations (e.g., Borghese, Minard, & Thoman, 1995) to the same outcomes at different ages. In line with Sadeh's (2003) suggestion that a developmental perspective is needed when assessing sleep problems in children, we argue that these age differences are a key aspect of the inconsistency in findings and that developmental considerations should guide the choice of sleep parameters likely to represent meaningful individual differences.

A first consideration is that while greater sleep duration is generally regarded as a positive parameter, a primary regulation task in infancy is to spend gradually *less* time asleep. As reviewed earlier, newborns spend about two thirds of each circadian period asleep, and sleep duration decreases gradually. Acebo et al. (2005) found this decline to be very clear during the 2nd year of life: They observed a significant decrease in total sleep duration reported by parents through a sleep diary between 1 and 2 years of age, followed by stabilization after 2 years. Greater sleep duration may thus be a developmentally inadequate index of sleep quality among infants, and in fact, total sleep duration (recorded by actigraphy) has been shown to be unrelated to cognitive performance in infants (Scher, 2005) and older children (Buckhalt et al., 2007; Sadeh et al., 2002). Another key developmental task in infancy is to adapt to the day–night alternation. Hence, while newborns sleep roughly equal amounts of time in daytime and at night, a circadian rhythm is established over the course of the 1st year, such that infants gradually get most of their sleep at night (Anders & Keener, 1985; Halpern, MacLean, & Baumeister, 1995), although important individual differences exist in the rate of adaptation to the day–night cycle (e.g., St. James-Roberts & Plewis, 1996). This adaptation mostly manifests itself through a decrease in daytime sleep. For instance, Acebo et al. (2005) found significant decreases in parent-reported number of daily naps and number of nap sleep minutes between 12 and 18 and between 18 and 24 months, which accounted for the overall decline in total sleep reported earlier. Hence, a greater proportion of total sleep occurring at night can be considered to be a marker of advanced biobehavioral organization. A third aspect of biological maturation in infancy is the consolidation of sleep patterns into longer uninterrupted bouts of sleep (Anders & Keener, 1985; Coons & Guilleminault, 1984), which also shows substantial interindividual variability (Sadeh & Anders, 1993). In fact, although the developmental milestone of "sleeping through the night" is most often achieved during the 1st year of life, as many as 20%–30% of infants continue to experience fragmented sleep past their 1st year (see Sadeh, 1996). Sleep fragmentation at night has been found to relate negatively to cognitive performance in infants (Scher, 2005) and children (Buckhalt et al., 2007; Sadeh et al., 2002).

This study set out to examine the prospective links between infant sleep regulation and subsequent EF in typically developing children, using a developmental approach to the assessment of both sleep and EF. The sleep parameters retained for their developmental relevance were: (a) the proportion of night time sleep to total sleep and (b) sleep fragmentation. Total sleep duration was also considered, given its frequent use as a sleep index. We used Carlson's (2005) empirically derived guidelines for the measurement of EF with toddlers. It was expected that a higher percentage of night time sleep and fewer nights awakenings at 12 and 18 months would relate to better EF at 18 and 26 months, but not to measures of overall cognitive functioning. In a comprehensive review of the literature on different components of EF among children suffering from sleep-disordered breathing, Beebe (2006) outlines that deficits are particularly consistent in the spheres of sustained attention, vigilance, and impulse control, whereas links with other executive functions, such as working memory or mental flexibility, are very limited. Sadeh et al. (2002) also reported specific links between school-age children's sleep fragmentation and impairments in impulse control. We thus expected that links between sleep and EF would be more marked in the sphere of impulse control.

Method

Participants

Sixty mother–infant dyads (36 girls and 24 boys) living in a large Canadian metropolitan area participated in this study. Families were recruited from birth lists provided by the Ministry of Health and Social Services. Criteria for participation were full-term pregnancy and the absence of any known physical or mental disability in the infant. Family income varied from < $20,000 CDN to over $100,000 CDN, with an average of $70,000 CDN. Mothers were between 24 and 45 years old ($M = 31.7$). Most (68.3%) had a college degree and 86.7% were Caucasian. All but 5 were married or living with the child's father throughout data collection.

Procedure

The dyads took part in three assessments, when children were 12–13 (T1; $M = 12.9$, $SD = 1.2$), 18 (T2; $M = 18.3$, $SD = 1.1$), and 26 months of age (T3; $M = 26.3$, $SD = 0.8$). Children's sleep regulation was assessed with a sleep diary completed by mothers at T1 and T2, whereas EF was assessed at T2 and T3 (see the following). Child overall cognitive functioning was assessed at all three time points (i.e., mental development at T1 and expressive vocabulary at T2 and T3).

Measures

Sleep Regulation

The parent sleep diary is a noninvasive measure widely used in sleep research with infants and children (e.g., Acebo et al., 2005; Corkum, Panton, Ironside, MacPherson, & Williams, 2008; Sadeh, 1994, 1996; Werner, Molinari, Guyer, & Jenni, 2008). The diary records, on a timeline of 24 hr for several consecutive days, the sleep pattern of the child as it unfolds. The parent is asked to indicate, for each half-hour, whether the child was awake or asleep, thus representing all periods of sleep and wake, including diurnal naps and nocturnal wakes. The parent is also requested to report any unusual event that might have disturbed the child's sleep pattern, such as illness or traveling. Although more days of assessment provide more data and thus can conceivably reduce measurement error, parent compliance in filling out the diary becomes an increasingly greater concern with lengthier periods of diary assessment (Sadeh, 2008). St. James-Roberts and Plewis (1996) found that 3 days represented the optimal duration to obtain reliable data while accounting for mothers' compliance in filling out diaries of infant fussing, feeding, and sleeping patterns. Consequently, mothers in this study were asked to complete the diary on 3 consecutive days during which their child had a fairly usual routine, at both 12 and 18 months of age. Three parameters were derived from the sleep diary: total sleep duration, percentage of total sleep occurring at night time (between 7 p.m. and 7 a.m.), and sleep fragmentation, that is, number of night awakenings. To account for different sleep schedules among children, wake periods occurring toward the beginning or the end of the predefined night period were considered night awakenings only if they were preceded and followed by at least 30 min of consecutive sleep.

Owing to some families' failure to return sleep diaries, or to diaries considered unusable (generally because they were completed on days marked by unusual events, such as sleeping at friends' homes or an illness in the child), 9 of the 60 families did not have valid sleep data at 18 months. Analyses pertaining to 18-month sleep are thus conducted with 51 children, whose mean age, other demographic information, and EF did not differ from the total sample.

Child EF: 18-Month Assessment

Until recently, the earliest measurement guidelines for EF started at 24 months (Carlson, 2005). However, based on neuropsychological testing, which suggests that certain aspects of EF begin to develop earlier, notably working memory, our team developed a downward adaptation of a working memory task originally designed for 2-year-old children.

Hide the Pots (adapted from Hughes & Ensor, 2005; see also Uzgiris & Hunt, 1975—Scale I, Search for Hidden

Objects). An attractive sticker was hidden in full sight of the child under one of three opaque pots of different colors. In the warm-up phase, no further action took place, and the child was immediately asked to retrieve the sticker where it was hidden. Three practice trials were conducted (one with each pot). This phase was intended to introduce children to what was expected. In the second phase, the experimenter also hid the sticker once under each pot but covered the pots with a blanket on each trial before immediately inviting the child to find the sticker. The child thus had to hold the location of the sticker in memory, remove the blanket, and reach for the pot where the sticker was hidden. Scores consisted of the number of trials (0–3) where the child found the sticker on his or her first attempt. In a prior study, this working memory task was shown to be predictive of later EF performance on well-validated tasks (Bernier, Carlson, & Whipple, 2010).

Child EF: 26-Month Assessment

The tasks conducted at 26 months were chosen based on Carlson's (2005) measurement guidelines with the aim of maximizing detection of individual differences in three dimensions of EF: working memory, inhibitory control, and set shifting.

Spin the Pots (Hughes & Ensor, 2005). Children were asked to search for stickers that were hidden in opaque pots, subsequently covered and rotated 180°. Six stickers were hidden in eight pots of very different visual appearances, thus leaving two pots empty. Each time a sticker was found, it was given to the child and all eight pots were covered and rotated 180° again. Therefore, as children found stickers, there was an increasingly large number of empty pots on the tray. The score was calculated as 16 (which is the maximum number of trials allowed) minus the number of errors made (i.e., looking under a pot in which no sticker had been hidden, or perseverative looking).

Delay of Gratification (Kochanska et al., 2000). The experimenter placed a present under a transparent cup and asked children to wait until she rang a bell before retrieving it. Four trials were conducted, where the child had to wait 5, 10, 15, and then 20 s. Scores were the number of seconds waited on each trial, that is, the time elapsed between the experimenter placing the cup and the child touching it (if he or she did so before maximum waiting time had elapsed).

Shape Stroop (Kochanska et al., 2000). Children were first shown six cards depicting three small and three large fruits, and were asked to point to each in turn to ensure that they knew the names and sizes of the fruits (e.g., ''Show me the big apple''). They were then shown three cards, each depicting one of the small fruits embedded in one of the larger (mismatching) ones, and asked to point to each of the small fruits in turn (e.g., ''Show me the *little* banana''). The score consisted of the number (0–3) of small fruits correctly pointed to.

Baby Stroop (adapted from Hughes & Ensor, 2005). Children learned a rule for feeding two dolls, feeding the ''mommy'' doll with a larger spoon and the baby doll with a smaller spoon. As soon as the child clearly understood the rule, it was reversed such that the child was asked to feed the larger doll with the smaller spoon once, and then to feed the baby doll with the larger spoon on the next trial. Children got 1 point for feeding each doll with the correct spoon on their first attempt. Scores thus ranged from 0 to 2.

At both T2 and T3, administration of the EF tasks was videotaped for later coding. Because we were interested in individual differences rather than group patterns, EF tasks were administered in fixed order.

General Cognitive Ability

At 12 months, children's general cognitive functioning was assessed with the Mental Development Index of the Bayley Scales of Infant Development (MDI; Bayley, 1993), a standardized test of cognitive development for children aged 1–42 months. A total of 23 protocols (38%) were double-coded by independent raters (using the video recording of the administration). Intraclass correlation (ICC) between raters' MDI scores suggested good reliability, ICC = .85.

At both 18 and 26 months, we assessed child expressive vocabulary, often used as a proxy for overall cognitive ability, with the MacArthur Communicative Development Inventory (MCDI; Fenson et al., 1994). Parents were asked to identify from a list which words they had heard their child say. Fenson et al. report excellent reliability indices for the instrument.

Results

Preliminary Analyses

Sleep Parameters

Stability of sleep parameters across the 3 days was moderate to high, with average between-day correlations of $r = .61$, $p < .001$, for total sleep duration; $r = .64$, $p < .001$, for percent of night time sleep; $r = .69$, $p < .001$, for number of night awakenings at 12 months; $r = .60$, $p < .001$, for total sleep duration; $r = .63$, $p < .001$, for percent of night sleep; and $r = .78$, $p < .001$, for number of night awakenings at 18 months. To increase reliability of estimates, data were thus averaged across the 3 days for each sleep parameter. Correlations were lower when looking at continuity between 12 and 18 months: $r = .44$, $p < .01$, for total sleep duration; $r = .46$, $p < .01$, for percent of night sleep; and $r = .52$, $p < .01$, for number of night awakenings. Given these correlations and the data presented earlier suggesting the importance of an age-sensitive approach to the assessment of sleep among infants, the 12- and 18-month parameters were thus kept as distinct variables in the remainder of the analyses. Table 1 presents descriptive statistics for all sleep and EF variables. The six sleep variables showed satisfactory variability, although night awakenings were fairly infrequently reported. EF tasks also showed adequate variability, although Baby Stroop was positively skewed, whereas the 5- and 10-s Delay of Gratification trials were negatively skewed.

Table 1
Mean, Standard Deviation, and Range for All Sleep and Executive Functioning (EF) Measures

Variable	*M*	*SD*	Observed range	Theoretical range
Sleep 12 months				
Total sleep duration (hours)	13.06	1.39	9.75–16.50	0–24
% of night sleep	78.20	5.80	61–91	0–100
Number of night awakenings	0.98	1.38	0–7	0–N/A
Sleep 18 months				
Total sleep duration (hours)	13.09	1.06	10.10–15.25	0–24
% of night sleep	79.40	6.40	57–91	0–100
Number of night awakenings	0.58	1.20	0–5	0–N/A
Child EF 18 months				
Hide the Pots	1.33	0.93	0–3	0–3
Child EF 26 months				
5-s delay	4.81	0.71	1–5	0–5
10-s delay	9.34	1.84	1–10	0–10
15-s delay	12.79	4.51	0–15	0–15
20-s delay	14.91	7.48	0–20	0–20
Spin the Pots	10.39	3.45	4–16	0–16
Shape Stroop	1.58	1.20	0–3	0–3
Baby Stroop	0.43	0.74	0–2	0–2

Note. Values for all sleep variables are averages of the 3 days of data.

Executive Function Data Reduction

The 26-month EF scores of all 60 children, including those for whom sleep data were not available at T2, were submitted to a principal component analysis (PCA) to reduce the number of data points and compute more reliable aggregate estimates. The PCA yielded a two-factor solution (eigenvalues > 1.0) representing 68.8% of the total variance. These two factors were submitted to a principal axis rotation (oblimin). Factor loadings for 15-s delay (.82), 10-s delay (.80), 5-s delay (.78), and 20-s delay (.56) suggest that the first factor taps impulse control, whereas the second factor appears to represent working memory, set shifting, and inhibitory control: Spin the Pots (.78), Baby Stroop (.58), and Shape Stroop (.37). No cross-loadings (above .25) were observed. Studies of EF in young children have found similar factor structures, in which tasks involving delay of gratification load on one factor, whereas tasks entailing different degrees of working memory, set shifting, and inhibitory control load on a separate factor (Carlson, Mandell, & Williams, 2004; Carlson & Moses, 2001; Davis-Unger & Carlson, 2008). The authors have labeled these two factors Delay (referred to here as Impulse Control, to highlight the skill required rather than the task) and Conflict-EF, respectively. Given that the current factor structure was very clear empirically and reproduced these two dimensions, two averaged standardized scores were computed and used in further analyses. Impulse Control and Conflict-EF were moderately correlated, $r = .33$, $p < .05$. The 18-month EF assessment consisted of the single task presented earlier and assessing working memory. This 18-month working memory score was correlated with 26-month Conflict-EF, $r = .32$, $p < .05$, and 26-month Impulse Control, $r = .27$, $p < .05$.

Sociodemographic Variables

We next examined whether sociodemographic variables (child age and gender, number of siblings, maternal and paternal age and education, and family income) were related to sleep or EF variables. Results showed that children of more educated mothers, $r = .23$, $p < .05$, and fathers, $r = .25$, $p < .05$, performed better on working memory at 18 months, and children of more educated mothers performed better on Conflict-EF at 26 months, $r = .23$, $p < .05$. Furthermore, children whose parents earned higher incomes had greater percentages of night sleep at 12 months, $r = .32$, $p < .01$, and performed better on Conflict-EF at 26 months, $r = .30$, $p < .05$. In contrast, no relations with EF or sleep were found for child gender, age, or number of siblings. Given the high correlations (ranging from .55 to .66) between maternal and paternal education and family income, these three variables were standardized and averaged into a composite index of socioeconomic status (SES), which is covaried in the main analyses.

General Cognitive Ability

Next, we examined whether child overall cognitive functioning was related to EF performance. As is often reported in previous studies (e.g., Carlson et al., 2004; Hughes & Ensor, 2005), prior mental development (12 months) was related to all three EF indices (*r*s varying between .22 and .27, *p*s < .05). Concurrent (18 or 26 months) expressive verbal ability was also associated with EF performance, with correlations varying between .30 and .34, *p*s < .05.

Last, correlations were run between indicators of overall cognitive functioning (mental development and verbal ability) and sleep parameters at 12 and 18 months. Only one correlation reached significance: A greater percentage of night time sleep at 12 months was related to a larger expressive vocabulary at 26 months, $r = .29$, $p < .05$. All other correlations were nonsignificant and below .17. Note that this pattern of relations was very similar when holding SES constant.

Table 2
Zero-Order Correlations Between Executive Functioning (EF) at 18 and 26 Months and Sleep Duration, Percent of Night Sleep, and Sleep Fragmentation at 12 and 18 Months

	18-month working memory	26-month conflict-EF	26-month impulse control
Sleep duration			
12 months	−.11	−.10	−.06
18 months	−.16	.09	−.16
% of night sleep			
12 months	.06	.27*	.32*
18 months	.35*	.13	.40**
Fragmentation			
12 months	.03	−.03	.04
18 months	.12	−.07	.01

*$p < .05$. **$p < .01$.

Main Analyses

To address our primary aim, we next examined whether sleep parameters were associated with concurrent and subsequent child performance on EF tasks. As displayed in Table 2, sleep duration and sleep fragmentation did not relate to any EF score, all *r*s < .16, *ns*. These two sleep indices were therefore dropped from further analyses. In contrast, percentages of night sleep were related to several EF indices. Percent of night sleep at 12 months was correlated with both Conflict-EF and Impulse Control at 26 months. Percent of night sleep at 18 months also correlated with Impulse Control at 26 months, although not with Conflict-EF. However, it was associated with concurrent working memory. In all cases, relations were positive; that is, children getting a greater proportion of their sleep at night performed better on concurrent or later EF tasks.

Thus, percent of night sleep at both ages considered was related to EF although not to general cognitive ability. To submit this to an especially stringent test focusing on what is unique to EF above overall cognitive functioning, each EF index was submitted to a hierarchical regression analysis involving three steps: SES was entered in a first block, followed by indicators of general cognitive performance in a second block, and by relevant sleep parameters in a third block. Owing to the correlations presented in Table 2, the third block consisted of the 18-month percent of night sleep when predicting 18-month working memory, and the 12-month percent of night sleep when predicting 26-month Conflict-EF. Both 12- and 18-month percents of night sleep were found to be uniquely related to 26-month Impulse Control (see Table 2), but their intercorrelation ($r = .46$, $p < .01$) argued against entering them in the same equation (predictors moderately to highly interrelated can cause unreliable estimates; Tabachnick & Fidell, 2007). Impulse Control was thus submitted to two distinct regression equations, one with each night sleep percent entered in the third block.

Table 3 (Equations 1–4) shows that SES predicted between 0.1% and 6.7% of the variance in EF performance, whereas overall cognitive performance contributed an additional 5.1%–10.2%, and sleep parameters added between 4.8% (Conflict-EF) and 9.8% (Impulse Control with 18-month night sleep) of explained variance. Specifically, after controlling for SES and cognitive performance, children getting a higher percentage of their sleep during the night at 12 months performed marginally better on

Table 3
Summary of Regression Analyses Predicting Executive Functioning (EF) Performance at 18 and 26 Months

	Criterion			
Model	R^2	ΔR^2	*F* change	β
1. 18-month working memory				
1. SES	.049		2.98†	.21†
2. Mental development				.35*
Verbal ability	.151	.102	3.52*	.11
3. 18-month % night sleep	.238	.087	4.51*	.29*
2. 26-month conflict-EF				
1. SES	.067		4.12*	.31*
2. Mental development				.25*
Verbal ability	.128	.061	3.12†	.13
3. 12-month % night sleep	.176	.048	3.13†	.22†
3. 26-month impulse control				
1. SES	.001		0.39	.01
2. Mental development				.10
Verbal ability	.052	.051	1.33	.17
3. 12-month % night sleep	.138	.086	4.78*	.30*
4. 26-month impulse control				
1. SES	.001		0.39	.01
2. Mental development				.10
Verbal ability	.052	.051	1.33	.17
3. 18-Month % night sleep	.150	.098	5.02*	.33*
5. 26-month conflict-EF				
1. SES	.067		4.12*	.31*
2. Mental development				.25*
Verbal ability	.128	.061	3.12†	.13
3. 18-month working memory	.170	.042	3.01†	.19†
4. 12-month % night sleep	.194	.024	1.21	.10
6. 26-month impulse control				
1. SES	.001		0.39	.01
2. Mental development				.10
Verbal ability	.052	.051	1.33	.17
3. 18-month working memory	.098	.046	2.98†	.18†
4. 12-month % night sleep	.151	.053	4.71*	.26*
7. 26-month impulse control				
1. SES	.001		0.39	.01
2. Mental development				.10
Verbal ability	.052	.051	1.33	.17
3. 18-month working memory	.098	.046	2.98†	.18†
4. 18-month % night sleep	.168	.070	4.99*	.30*

Note. SES = socioeconomic status.
$^{\dagger}p < .10$. $^{*}p < .05$.

Conflict-EF, β = .22, $p < .10$, and significantly better on Impulse Control at 26 months, β = .30, $p < .05$. In addition, children getting a higher percentage of their sleep during the night at 18 months performed better on concurrent working memory, β = .29, $p < .05$, and later Impulse Control, β = .33, $p < .05$. Hence, all direct relations uncovered between sleep and EF (Table 2) remained significant after accounting for SES and general cognitive ability, with the exception of the relation between 12-month night sleep percent and 26-month Conflict-EF, which became marginally significant.

Finally, we examined whether sleep was related to later changes in EF performance by analyzing relations between night sleep percents and 26-month EF, partialing out previous (18-month) working memory, in addition to the previous controls. Working memory at 18 months was therefore included in a third block of the three equations predicting 26-month EF, whereas sleep parameters were entered last, in a fourth block of each equation. Results, presented in Table 3 (Equations 5–7), indicated that 26-month Impulse Control was positively associated with night sleep percent at both 12, β = .26, $p < .05$, and 18 months, β = .30, $p < .05$, over and above SES, general cognitive ability, and prior EF performance.

Discussion

The aim of this report was to investigate the prospective links between sleep regulation and EF in normally developing infants and toddlers. Based on prior work with normative and clinical populations of older children, it was expected that more mature sleep, indicated by a higher percentage of total sleep occurring during night time and fewer night awakenings, would relate to better performance on EF tasks but not to general cognitive functioning. With some exceptions, the results provided support for this hypothesis. Whereas only one relation between sleep and general cognitive functioning (i.e., 26-month vocabulary) was found to be significant, the percentage of night time sleep, at both 12 and 18 months of age, was associated with several EF indices. Hence, 12-month night sleep was related to both conflict-EF and impulse control at 26 months of age, whereas 18-month night sleep was related to concurrent working memory and later impulse control. Children getting a higher percentage of their sleep during night time were more advanced in their EF development. Importantly, these relations appear to be robust and specific to EF, as they held after controlling for SES as well as for two well-documented correlates of EF: overall mental development and verbal ability. Finally, the longitudinal links between percentage of night time sleep at 12 and 18 months and 26-month impulse control both remained significant even after controlling for 18-month working memory, SES, Bayley's MDI, and MacArthur verbal ability scores. In contrast, parent-reported night awakenings at either age did not relate to EF performance, nor did total sleep duration.

These findings provide support for the hypothesis that sleep is specifically associated with frontal/executive functions in children. The current results are in line with previous correlational and experimental studies with adults (see Dahl, 1996; Harrison & Horne, 1998; Jones & Harrison, 2001) and children of clinical (see Beebe, 2006) and normative populations (Randazzo et al., 1998; Sadeh et al., 2002, 2003). They also extend previous findings in several important ways. Notably, this is the first study, to our knowledge, to use a longitudinal design and developmentally sensitive measures to investigate prospective links between early sleep and later child EF. Such an approach is especially helpful in disentangling temporary child sleepiness from enduring effects of impaired sleep. In addition, previous studies investigated sleep–EF links in children 7 years and older, that is, after the decline in neural density of the frontal lobes has begun (Huttenlocher, 2002). Finding similar relations between sleep and performance on frontal tasks much earlier in life speaks to the developmental robustness of the associations. We see these results as particularly meaningful given that they are based on naturally occurring variations in sleep patterns, which are much less dramatic than diagnosed sleep disorders or experimental manipulations of sleep duration.

These results raise the possibility that infant sleep affects developing brain structures in the first 2 years of life, thereby setting in motion a cascade of neural effects that may carry substantial implications for later EF. The multiple studies that have demonstrated that sleep plays a crucial role in learning and memory among adults (see Rauchs, Desgranges, Foret, & Eustache, 2005; Walker & Stickgold, 2006) are useful in identifying the mechanisms through which sleep could affect brain development. It is increasingly believed that a number of cellular and biochemical mechanisms during sleep may facilitate brain plasticity. Animal and human studies have shown that neuronal circuits active during previous learning are reactivated during sleep (Peigneux et al., 2004; Rasch & Born, 2007). In addition, proteins associated with brain plasticity and learning are up-regulated during sleep (Basheer, Brown, Ramesh, Begum, & McCarley, 2005; Cirelli, 2005). Studies in kittens have demonstrated a direct impact of sleep and sleep deprivation on experience-dependent brain plasticity (Frank, Issa, & Stryker, 2001; Frank, Jha, & Coleman, 2006; Jha et al., 2005). It has also been proposed that sleep would induce a progressive downscaling of synaptic strength accumulated during the day to reach an energetically suitable level and make efficient use of cerebral gray matter (Tononi & Cirelli, 2006). Finally, Gaillot (2008) cited evidence showing that brain glycogen, a source of stored energy that is used to support brain activity, is depleted during activities that tax executive control systems and restored during sleep. He thus proposed that one function of sleep might be to replenish brain glycogen stores, which in turn accounts for improved executive function performance. These restorative features of sleep may be especially prominent in the frontal cortex, which is one of the most active and most densely connected brain regions during wake time, and may thus require more recuperation during sleep (Cajochen et al., 2001; Dahl, 1996; Horne, 1993). The findings of this study are consistent with the notion of a direct impact of sleep on brain plasticity, and thus brain development, beginning early in life.

In contrast, our results could also suggest that both percentage of night time sleep and performance on EF tasks are indices of maturation of the central nervous system, in the absence of a causal link between the two. The decline in total sleep duration, largely attributable to a decrease in daytime sleep (Acebo et al., 2005; Coons & Guilleminault, 1992; Davis et al., 2004; Iglowstein et al., 2003), and adaptation to the day–night alternation, resulting in increased night sleep and decreased daytime sleep (Anders & Keener, 1985; Halpern et al., 1995; St. James-Roberts & Plewis, 1996) are well-documented changes in the organization of sleep–wake patterns during infancy, which are often presumed to result from neurological maturation. In the same vein, maturation of the frontal lobes is associated with increments in EF performance during early childhood (see Zelazo et al., 2008). Hence, although the links observed here between sleep and EF appear robust, further research is needed to sort out the sequence of events. Adult and animal brain plasticity research suggests that infant sleep could affect the developing brain structures directly through a number of biochemical mechanisms identified earlier, and experimental studies with both children and adults do suggest a causal link between sleep and EF performance (e.g., Jones & Harrison, 2001; Sadeh et al., 2003). However, the current correlational design does not rule out an alternative explanation, according to which brain maturation would be reflected in both the early organization of sleep–wake patterns and later EF, rather than being the mechanism accounting for a causal link between early sleep and later EF. In fact, the two possibilities are not mutually exclusive, and we would argue that dynamic bidirectional effects are likely to be at play, in which initial maturation of brain structures allows for the establishment of an increasingly mature sleep–wake organization, which in turn favors further brain maturation/plasticity and hence, the emergence of executive control capacities.

An alternate account of the findings involves the family context, which research increasingly suggests is closely linked to child sleep (see Dahl & El-Sheikh, 2007; Jenni & O'Connor, 2005). Factors, such as SES (Acebo et al., 2005), family stress (e.g., Bates, Viken, Alexander, Beyers, & Stockton, 2002), parental psychopathology (Seifer, Sameroff, Dickstein, Hayden, & Schiller, 1996), and marital discord (El-Sheikh, Buckhalt, Mize, & Acebo, 2006) have all been shown to be linked to children's sleep. Specific parental behaviors have been shown to relate to child sleep as well, notably parental warmth (Adam, Snell, & Pendry, 2007) and bedtime behaviors (Morrell & Steele, 2003). Importantly, family factors, such as SES (e.g., Mezzacappa, 2004) and parental behaviors (Bernier et al., 2010; Hughes & Ensor, 2009) are increasingly documented correlates of child EF as well. Hence, one may propose that the links found here between child sleep and EF could be owing to a common underlying influence of the family environment: Well-organized families, characterized by clear limit setting, and warm parent–child relations, could constitute a context likely to favor the orderly development of both sleep patterns and EF in the young child. However, the links between child sleep and family functioning appear to be bidirectional, as infant and child sleep difficulties have been found to predict parents' own sleep difficulties, marital discord, negative mood, and increased stress (Meijer & van den Witenboer, 2007; Meltzer & Mindell, 2007). This raises yet another hypothesis: that early sleep is linked to child EF through its intermediate impact on parents' psychological adjustment. Hence, while the current findings add to an emerging literature documenting links between child sleep and EF, longitudinal designs entailing careful examination of family context variables are necessary for the continued investigation of the neural and social mechanisms linking sleep, EF, and family relations.

Overall, the results suggest that the proportion of night time sleep at 12 and 18 months is relevant to the prediction of child EF. Results were different when considering the other sleep parameters. Although the absence of relations between EF and total sleep duration was expected because of the developmentally inappropriate nature of sleep duration as an index of sleep maturation during infancy, we did expect to find relations with night awakenings. The null findings for sleep fragmentation, rather than questioning the importance of long uninterrupted bouts of sleep in early EF development, may reflect methodological imprecision. In contrast to general sleep questionnaires, sleep diary data converge with objective sleep data gathered by actigraphy, including studies of infants and children (Acebo et al., 2005; Sadeh, 1996, 2004; Werner et al., 2008). In fact, a parent sleep diary is proposed to be interchangeable with actigraphy when assessing sleep schedules, for instance, sleep onset and duration (Sadeh, 2008; Werner et al., 2008). However, the diary method yields less individual variation than actigraphy (Sadeh, 1996), particularly with respect to night awakenings, which tend to be underestimated by parents (Sadeh, 2008; Werner et al., 2008). For instance, Sadeh (1996) found a correlation of .60 between night awakenings reported by parents and as assessed by actigraphy among infants 7–26 months old, while correlations for sleep schedule parameters (e.g., sleep duration) were all above .70. This would be because of many infants' capacity to self-soothe and go back to sleep without parental intervention. In line with this, Keener, Zeanah, and Anders (1988) as well as Minde et al. (1993) observed that while a majority of infants did wake up during the night, a substantial proportion could fall back asleep by themselves, and their awakenings thus went unnoticed by their parents. Accordingly, the use of objective sleep assessments, such as actigraphy, appears necessary for a thorough investigation of the links between early EF development and more subtle indices of sleep consolidation, especially night awakenings.

Another limitation of this study is that given the modest sample size, we could not correct statistically for the number of analyses run without compromising statistical power. This may have increased the risk of Type I error, despite the steps that were undertaken to reduce the number of analyses and variables. Future studies should therefore confirm the current findings with larger samples to achieve an optimal balance between Type I and Type II errors.

The longitudinal findings with the impulse control aspect of EF at 26 months were more compelling and more robust statistically than those with conflict-EF. Impulse control was related to the percentage of night time sleep assessed both 8 and 14 months earlier, and these relations held above

SES, general cognitive functioning, and even 18-month EF. These findings pointing to impulse control as particularly sensitive to sleep are in line with the results found in pediatric populations suffering from sleep apnea (for a review, see Beebe, 2006), with Sadeh et al.'s (2002) findings with nonclinical school-age children, and are reminiscent of the sleep difficulties often reported among children suffering from externalizing disorders characterized by poor impulse control (Owens, 2005; Touchette et al., 2007). They also remind us that EF is constituted of a heterogeneous set of skills (see Garon et al., 2008) that may show distinct relations to different aspects of the child's experience. It was recently found that conflict-EF is related to the quality of prior parent–infant interactions, whereas impulse control is not (Bernier et al., in press). One could thus speculate that different experiential pathways lead to the development of specific aspects of EF, with perhaps social and relational factors (e.g., language stimulation and parent scaffolding) playing a greater role in conflict-EF, whereas impulse control may be driven to a greater degree by biological factors, such as sleep, temperament, or neuroendocrine regulation. One should bear in mind, however, that these putative pathways would likely not be independent of one another, as more organized and well-functioning families appear better equipped to foster both infant sleep regulation and the development of EF.

This study suggested that greater proportions of total sleep occurring at night time are related to children's higher EF performance up to 14 months later, especially on tasks involving a strong impulse control component. These findings add to an emerging literature focusing on the developmental antecedents of children's EF, which has begun to identify parenting as a social precursor of child EF (Bernier et al., 2010; Hughes & Ensor, 2009). Given that parenting is proposed to play a role in the consolidation of sleep–wake patterns (Anders, 1994; Goodlin-Jones, Burnham, Gaylor, & Anders, 2001), and that there is some overlap between different aspects of EF, it is likely that biological and environmental factors interact in complex ways in shaping EF, with specific biological and environmental factors carrying different weights in the prediction of different aspects of EF at different ages. Given the compelling research demonstrating that variation within the normative range of EF performance is meaningful for child adjustment in several spheres, an important agenda for future research is to unravel the neural and social pathways that lead to its optimal development.

References

Acebo, C., Sadeh, A., Seifer, R., Tzischinsky, O., Hafer, A., & Carskadon, M. A. (2005). Sleep/wake patterns derived from activity monitoring and maternal report for healthy 1- to 5-year-old children. *Sleep, 28*, 1568–1577.

Achermann, P., Dijk, D. J., Brunner, D. P., & Borbély, A. (1993). A model of human sleep homeostasis based on EEG slow-wave activity: Quantitative comparison of data and simulations. *Brain Research Bulletin, 31*, 97–113.

Adam, E. K., Snell, E. K., & Pendry, P. (2007). Sleep timing and quantity in ecological and family context: A nationally representative time-diary study. *Journal of Family Psychology, 21*, 4–19.

Anders, T. F. (1994). Infant sleep, nighttime relationships, and attachment. *Psychiatry, 57*, 11–21.

Anders, T. F., & Keener, M. (1985). Developmental course of nighttime sleep-wake patterns in full-term and premature infants during the first year of life. *Sleep, 8*, 173–192.

Anders, T. F., Keener, M. A., & Kraemer, H. (1985). Sleep-wake state organization, neonatal assessment and development in premature infants during the first year of life. *Sleep, 8*, 193–206.

Basheer, R., Brown, R., Ramesh, V., Begum, S., & McCarley, R. W. (2005). Sleep deprivation-induced protein changes in basal forebrain: Implications for synaptic plasticity. *Journal of Neuroscience Research, 82*, 650–658.

Bates, J. E., Viken, R. J., Alexander, D. B., Beyers, J., & Stockton, L. (2002). Sleep and adjustment in preschool children: Sleep diary reports by mothers relate to behavior reports by teachers. *Child Development, 73*, 62–74.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). New York: PsychCorp.

Becker, P. T., & Thoman, E. B. (1981). Rapid eye movement storms in infants: Rate of occurrence at 6 months predicts mental development at 1 year. *Science, 212*, 1415–1416.

Beebe, D. W. (2006). Neurobehavioral morbidity associated with disordered breathing during sleep in children: A comprehensive review. *Sleep, 29*, 1115–1134.

Beebe, D. W., & Gozal, D. (2002). Obstructive sleep apnea and the prefrontal cortex: Towards a comprehensive model linking nocturnal upper airway obstruction to daytime cognitive and behavioral deficits. *Journal of Sleep Research, 11*, 1–16.

Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development 81*, 326–339.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647–663.

Boonstra, T. W., Stins, J. F., Daffertshofer, A., & Beek, P. J. (2007). Effects of sleep deprivation on neural functioning: An integrative review. *Cellular & Molecular Life Sciences, 64*, 934–946.

Borghese, I. F., Minard, K. L., & Thoman, E. B. (1995). Sleep rhythmicity in premature infants: Implications for development status. *Sleep, 18*, 523–530.

Buckhalt, J. A., El-Sheikh, M., & Keller, P. (2007). Children's sleep and cognitive functioning: Race and socioeconomic status as moderators of effects. *Child Development, 78*, 213–231.

Cajochen, C., Foy, R., & Dijk, D. J. (1999). Frontal predominance of a relative increase in sleep delta and theta EEG activity after sleep loss in humans. *Sleep Research Online, 2*, 65–69.

Cajochen, C., Knoblauch, V., Krauchi, K., Renz, C., & Wirz-Justice, A. (2001). Dynamics of frontal EEG activity, sleepiness and body temperature under high and low sleep pressure. *NeuroReport, 12*, 2277–2281.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–616.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40*, 1105–1122.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22*, 489–510.

Cirelli, C. (2005). A molecular window on sleep: Changes in gene expression between sleep and wakefulness. *Neuroscientist, 11*, 63–74.

Coons, S., & Guilleminault, C. (1984). Development of consolidated sleep and wakeful periods in relation to the day/night cycle in infancy. *Developmental Medicine and Child Neurology, 26*, 169–176.

Coons, S., & Guilleminault, C. (1992). Development of sleep-wake patterns and REM sleep stages during the first six months of life in normal infants. *Pediatrics, 69*, 793–798.

Corkum, P., Panton, R., Ironside, S., MacPherson, M., & Williams, T. (2008). Acute impact of immediate release methylphenidate administered three times a day on sleep in children with attention-deficit/hyperactivity disorder. *Journal of Pediatric Psychology, 33*, 368–379.

Dahl, R. E. (1996). The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology, 8*, 3–27.

Dahl, R. E., & El-Sheikh, M. (2007). Considering sleep in a family context: Introduction to the special issue. *Journal of Family Psychology, 21*, 1–3.

Davis, K. F., Parker, K. P., & Montgomery, G. L. (2004). Sleep in infants and young children: Part one: Normal sleep. *Journal of Pediatric Health Care, 18*, 65–71.

Davis-Unger, A., & Carlson, S. M. (2008). Children's teaching: Relations to theory of mind and executive function. *Mind, Brain, and Education, 2*, 128–135.

Dearing, E., McCartney, K., Marshall, N. L., & Warner, R. M. (2001). Parental reports of children's sleep and wakefulness: Longitudinal associations with cognitive and language outcomes. *Infant Behavior and Development, 24*, 151–170.

El-Sheikh, M., Buckhalt, J. A., Mize, J., & Acebo, C. (2006). Marital conflict and disruption of children's sleep. *Child Development, 77*, 31–43.

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*, 465–486.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (5, Serial No. 242).

Frank, M. G., Issa, N. P., & Stryker, M. P. (2001). Sleep enhances plasticity in the developing visual cortex. *Neuron, 30*, 1–20.

Frank, M. G., Jha, S. K., & Coleman, T. (2006). Blockade of postsynaptic activity in sleep inhibits developmental plasticity in visual cortex. *NeuroReport, 17*, 1459–1463.

Gaillot, M. T. (2008). Unlocking the energy dynamics of executive functioning: Linking executive functioning to brain glycogen. *Perspectives on Psychological Science, 3*, 245–263.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Gertner, S., Greenbaum, C. W., Sadeh, A., Dolfin, Z., Sirota, L., & Ben-Nun, Y. (2002). Sleep-wake patterns in preterm infants and 6 month's home environment: Implications for early cognitive development. *Early Human Development, 68*, 93–102.

Gomez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*, 670–674.

Goodlin-Jones, B. L., Burnham, M. M., Gaylor, E. E., & Anders, T. F. (2001). Night waking, sleep-wake organization, and self-soothing in the first year of life. *Journal of Developmental and Behavioral Pediatrics, 22*, 226–233.

Halpern, L. F., MacLean, W. E., & Baumeister, A. A. (1995). Infant sleep-wake characteristics: Relation to neurological status and the prediction of developmental outcome. *Developmental Review, 15*, 255–291.

Harrison, Y., & Horne, J. A. (1998). Sleep loss impairs short and novel language tasks having a prefrontal focus. *Journal of Sleep Research, 7*, 95–100.

Hayes, M. (2002). Methodological issues in the study of arousal and awakenings during sleep in the human infants. In P. Salzarulo & G. Ficca (Eds.), *Awakening and sleep-wake cycle across development* (pp. 21–45). Amsterdam: John Benjamins.

Horne, J. A. (1993). Human sleep, sleep loss and behaviour implications for the prefrontal cortex and psychiatric disorder. *British Journal of Psychiatry, 162*, 413–419.

Hughes, C. H., & Ensor, R. A. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology, 28*, 645–668.

Hughes, C. H., & Ensor, R. A. (2009). How do families help or hinder the emergence of early executive function? In C. Lewis & J. I. M. Carpendale (Eds.), *New directions in child and adolescent psychiatry* (Vol. 123, pp. 35–50). Hoboken, NJ: Wiley.

Huttenlocher, P. R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex.* Cambridge, MA: Harvard University Press.

Iglowstein, I., Jenni, O. G., Molinari, L., & Largo, R. H. (2003). Sleep duration from infancy to adolescence: Reference values and generational trends. *Pediatrics, 111,* 302–307.

Jenni, O. G., & Carskadon, M. A. (2007). Sleep behavior and sleep regulation from infancy through adolescence: Normative aspects. *Sleep Medicine Clinics, 2,* 321–329.

Jenni, O. G., & O'Connor, B. B. (2005). Children's sleep: An interplay between culture and biology. *Pediatrics, 115,* 204–216.

Jha, S. K., Jones, B. E., Coleman, T., Steinmetz, N., Law, C. T., Griffin, G., et al. (2005). Sleep-dependent plasticity requires cortical activity. *Journal of Neuroscience, 25,* 9266–9274.

Jones, B. E. (2005). From waking to sleeping: Neuronal and chemical substrates. *Trends in Pharmacological Sciences, 26,* 578–586.

Jones, K., & Harrison, Y. (2001). Frontal lobe function, sleep loss and fragmented sleep. *Sleep Medicine Reviews, 5,* 463–475.

Keener, M. A., Zeanah, C. H., & Anders, T. F. (1988). Infant temperament, sleep organization, and nighttime parental interventions. *Pediatrics, 81,* 762–771.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36,* 220–232.

Meijer, A. M., & van den Witenboer, G. L. H. (2007). Contribution of infants' sleep and crying to marital relationship of first-time parent couples in the 1st year after childbirth. *Journal of Family Psychology, 21,* 49–57.

Meltzer, L. J., & Mindell, J. A. (2007). Relationship between child sleep disturbances and maternal sleep, mood, and parenting stress: A pilot study. *Journal of Family Psychology, 21,* 67–73.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development, 75,* 1373–1386.

Minde, K., Popiel, K., Leos, N., Falkner, S., Parker, K., & Handley-Derry, M. (1993). The evaluation and treatment of sleep disturbances in young children. *Journal of Child Psychology and Psychiatry, 34,* 521–533.

Morrell, J., & Steele, H. (2003). The role of attachment security, temperament, maternal perception, and caregiving behavior in persistent infant sleeping problems. *Infant Mental Health Journal, 24,* 447–468.

Nelson, C. A., Thomas, K. M., & de Haan, M. (2006). Neural bases of cognitive development. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner. *Handbook of child psychology: Vol. 2. cognition, perception, and language* (6th ed., pp. 3–57). Hoboken, NJ: Wiley.

Owens, J. A. (2005). The ADHD and sleep conundrum: A review. *Journal of Developmental and Behavioral Pediatrics, 26,* 312–322.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., & Reggers, J. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron, 44,* 535–545.

Randazzo, A. C., Muehlbach, M. J., Schweitzer, P. K., & Walsh, J. K. (1998). Cognitive function following acute sleep restriction in children ages 10–14. *Sleep, 21,* 861–868.

Rasch, B., & Born, J. (2007). Maintaining memories by reactivation. *Current Opinion in Neurobiology, 17,* 698–703.

Rauchs, G., Desgranges, B., Foret, J., & Eustache, F. (2005). The relationships between memory systems and sleep stages. *Journal of Sleep Research, 14,* 123–140.

Sadeh, A. (1994). Assessment of intervention for infant night waking: Parental reports and activity-based home monitoring. *Journal of Consulting and Clinical Psychology, 62,* 63–68.

Sadeh, A. (1996). Evaluating night wakings in sleep disturbed infants: A methodological study of parental reports and actigraphy. *Sleep, 19,* 757–762.

Sadeh, A. (2003). Clinical assessment of pediatric sleep disorders. In M. L. Perlis & K. L. Lichstein (Eds.), *Treating sleep disorders: Principles and practice of behavioral sleep medicine* (pp. 344–364). Hoboken, NJ: Wiley.

Sadeh, A. (2004). A brief screening questionnaire for infant sleep problems: Validation and findings for an Internet sample. *Pediatrics, 113,* 570–577.

Sadeh, A. (2008). Commentary: Comparing actigraphy and parental report as measures of children's sleep. *Journal of Pediatric Psychology, 33,* 406–407.

Sadeh, A., & Anders, T. F. (1993). Infant sleep problems: Origins, assessment, interventions. *Infant Mental Health Journal, 14,* 17–34.

Sadeh, A., Gruber, R., & Raviv, A. (2002). Sleep, neurobehavioral functioning, and behavior problems in school-age children. *Child Development, 73,* 405–417.

Sadeh, A., Gruber, R., & Raviv, A. (2003). The effects of sleep restriction and extension on school-age children: What a difference an hour makes. *Child Development, 74,* 444–455.

Scher, A. (2005). Infant sleep at 10 months of age as a window to cognitive development. *Early Human Development, 81,* 289–292.

Seifer, R., Sameroff, A. J., Dickstein, S., Hayden, L. C., & Schiller, M. (1996). Parental psychopathology and sleep variation in children. *Child and Adolescent Psychiatric Clinics of North America, 5,* 715–727.

Singer, W. (1995). Development and plasticity of cortical processing architectures. *Science, 270,* 758–764.

St. James-Roberts, I., & Plewis, I. (1996). Individual differences, daily fluctuations, and developmental changes in amounts of infant waking, fussing, crying, feeding, and sleeping. *Child Development, 67*, 2527–2540.

Steenari, M. R., Vuontela, V., Paavonen, E. J., Carlson, S., Fjallberg, M., & Aronen, E. (2003). Working memory and sleep in 6- to 13-year-old schoolchildren. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 85–92.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Educational.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*, 49–62.

Touchette, É., Petit, D., Séguin, J. R., Boivin, M., Tremblay, R. E., & Montplaisir, J. Y. (2007). Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. *Sleep, 30*, 1213–1219.

Uzgiris, I. C., & Hunt, J. Mc. V. (1975). *Assessment in infancy: Ordinal scales of psychological development*. Champaign: University of Illinois Press.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57*, 139–166.

Werner, H., Molinari, L., Guyer, C., & Jenni, O. G. (2008). Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. *Archives of Pediatrics and Adolescent Medicine, 162*, 350–358.

Werth, E., Achermann, P., & Borbély, A. A. (1996). Brain topography of the human sleep EEG: Antero-posterior shifts of spectral power. *NeuroReport, 8*, 123–127.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 553–574). Cambridge, MA: MIT Press.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*, 198–226.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3, Serial No. 274).

Child Development, November/December 2010, Volume 81, Number 6, Pages 1753–1767

Pathways to Mathematics: Longitudinal Predictors of Performance

Jo-Anne LeFevre and Lisa Fast
Carleton University

Sheri-Lynn Skwarchuk
University of Winnipeg

Brenda L. Smith-Chant
Trent University

Jeffrey Bisanz
University of Alberta

Deepthi Kamawar and Marcie Penner-Wilger
Carleton University

A model of the relations among cognitive precursors, early numeracy skill, and mathematical outcomes was tested for 182 children from 4.5 to 7.5 years of age. The model integrates research from neuroimaging, clinical populations, and normal development in children and adults. It includes 3 precursor pathways: quantitative, linguistic, and spatial attention. These pathways (a) contributed independently to early numeracy skills during preschool and kindergarten and (b) related differentially to performance on a variety of mathematical outcomes 2 years later. The success of the model in accounting for performance highlights the need to understand the fundamental underlying skills that contribute to diverse forms of mathematical competence.

How do children acquire skill in conventional mathematics? Researchers have used information from a variety of sources to address this question, including longitudinal studies of normally developing children (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Bull, Espy, & Wiebe, 2008; DeSmedt, Verschaffel, & Ghesquière, 2009; Durand, Hulme, Larkin, & Snowling, 2005; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan, Kaplan, Oláh, & Locuniak, 2006; Locuniak & Jordan, 2008; Passolunghi, Vercelloni, & Schadee, 2007), research with children who have severe mathematics difficulties (i.e., dyscalculia; Berch & Mazzocco, 2007), and studies of children who have congenital syndromes that affect their mathematical learning (Ansari et al., 2003; Barnes, Smith-Chant, & Landry, 2005; DeSmedt et al., 2007; Mazzocco & McCloskey, 2005). However, only a few models of the relations between early cognitive abilities and mathematics outcomes have emerged (Bisanz, Sherman, Rasmussen, & Ho, 2005; cf. Desoete & Grégoire, 2006; Durand et al., 2005; Krajewski & Schneider, 2009). Furthermore, it is unclear how early numeracy skills (i.e., children's evolving knowledge about number and quantity) are connected to later mathematical performance. To address these gaps, we propose and test a model of mathematical development in which linguistic, quantitative, and spatial attention skills have distinct roles in children's acquisition of mathematics.

Proposed Model: Pathways to Mathematics

Dehaene and colleagues (Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, Piazza, Pinel, & Cohen, 2005; Spelke & Dehaene, 1999) have proposed that numerical processing involves three separate neural circuits in the parietal lobe. First, linguistic numerical tasks, such as retrieving answers to arithmetic combinations (e.g., 6 × 7), activate a region of the left angular gyrus that has been linked to language processing, including reading and phoneme

---

This research was supported by Standard Operating Grants from the Social Sciences and Humanities Research Council of Canada to J. LeFevre, J. Bisanz, S. Skwarchuk, B. Smith-Chant, and D. Kamawar. We thank the many graduate student colleagues and undergraduate researchers who contributed to this research. Lori Mergulhao in Winnipeg and Nicole Vachon in Peterborough provided invaluable organizational acumen. We thank all of the children who participated for their interest and enthusiasm. We are also grateful to the teachers, principal, parents, and other school board contacts who enabled this research.

Correspondence concerning this article should be addressed to Jo-Anne LeFevre, Institute of Cognitive Science, Carleton University, Ottawa, ON K1S 5B6, Canada. Electronic mail may be sent to jo-anne_lefevre@carleton.ca.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0011

detection. The second circuit activates portions of the posterior superior parietal lobule and is involved in spatial attention. The third circuit is activated in quantity tasks, such as magnitude comparison. It implicates the horizontal segment of the intraparietal sulcus (HIPS). Butterworth and colleagues (Butterworth, 1999, 2005; Castelli, Glaser, & Butterworth, 2006; Gelman & Butterworth, 2005) also argue for a brain-based number module in the HIPS that is specialized for processing numerosity or quantity. They propose that impairment of this module may be the source of developmental dyscalculia. Notably, although both Dehaene and Butterworth propose a ''core quantity system,'' Dehaene's model is focused on the processing of approximate magnitudes, whereas Butterworth proposes that the core quantity system is specialized to process small exact quantities. Despite these differences across theories, researchers generally agree that only the circuit associated with processing magnitude represents a distinct numerical component. The other two circuits involved in numerical tasks, the linguistic and spatial attentional circuits, are shared across a variety of cognitive tasks.

This neuropsychological account of numerical processing is consistent with the view that the development of mathematical skills involves several distinct cognitive systems (see also Ansari, 2008; Halberda, Mazzocco, & Feigenson, 2008). Accordingly, we formulated a model of numerical development that has three independent cognitive precursors or *pathways*: linguistic, quantitative, and spatial attention. All three pathways are hypothesized to contribute both to the development of early numeracy skills (i.e., before children have much formal instruction) and to children's acquisition of formal mathematical knowledge in school. Based on the model, we proposed two hypotheses about the course of numerical development in children between the ages of 4 and 8. In this article, we tested the model with longitudinal data.

Assessing the Pathways Model

Hypothesis 1: Linguistic, quantitative, and spatial attention pathways contribute independently to early numeracy

The first hypothesis was that linguistic, quantitative, and spatial attention precursor skills form distinct pathways that are independently related to early numeracy performance (i.e., knowledge that children acquire before formal schooling). This hypothesis leads to four predictions. First, linguistic skills should strongly predict children's performance on early numeracy measures that involve knowledge of the symbolic number system (such as naming numbers or writing Arabic digits) because learning the rules of the number system is assumed to be similar to mastering any symbolic representational system (e.g., written language). Second, quantitative precursors should predict early numeracy skills that require children to represent and operate on quantities (e.g., determining more and less, part–whole comparisons, and nonverbal arithmetic). Third, if the two skill pathways are distinct in the early stages of numeracy acquisition, then linguistic and quantitative early numeracy skills will not be strongly related. Fourth, spatial attention will contribute independently of linguistic and quantitative skills to both kinds of early numeracy performance. This prediction is consistent with Dehaene et al.'s (2005) view that the parietal circuit involved in spatial attention is general and not specifically connected to particular numerical tasks.

Children's early numeracy skills, measured before or during kindergarten, predict performance on conventional mathematics tests in the first few years of formal schooling (e.g., Aunola et al., 2004; Bull et al., 2008; Jordan et al., 2006; Jordan et al., 2007; Passolunghi et al., 2007). However, researchers have not identified a consistent and coherent set of cognitive skills that are predictive of early numeracy development. The Pathways model provides a framework to organize potential precursors. Aunio et al. (2006) reported support for distinct linguistic and quantitative pathways for Chinese and Finnish children, ranging in age from 4.5 to 7.5 years. Chinese children outperformed the Finnish children on the linguistic tasks (e.g., producing the number word sequence to 20) at both younger and older ages, but on the quantitative tasks (e.g., determining which set of objects had more) Chinese children outperformed the Finnish children only at the older ages. Similarly, Jordan et al. (2006) obtained a two-factor solution in an analysis of their early numeracy battery: Tasks requiring linguistic number knowledge (story problems and verbally presented number combinations) were distinguished from all other measures, most of which required quantitative knowledge but little knowledge of the formal number system. Jordan et al. (2007), following the same children from Jordan et al. (2006) into Grade 1, found differential patterns of growth for the linguistic and quantitative skills. With 7- to 10-year-old children, Durand et al. (2005) reported that both verbal ability and speeded magnitude comparison (e.g., Which is larger, 3 or 5?) predicted arithmetic performance. Krajewski and Schneider (2009) proposed a model that included distinct roles for linguistic knowledge and spatial attention in early numeracy development, although their model does not have a distinct quantitative pathway. In summary, existing research is generally consistent with the hypothesis that linguistic and quantitative pathways can be distinguished in early numeracy development.

Empirical support for a connection between spatial attention and mathematical learning comes from research on visual-spatial working memory. Working memory can be defined as the processing and storage requirements involved in complex cognitive tasks (Baddeley, 2002; Bull et al., 2008; DeStefano & LeFevre, 2004; Zentall, 2007). Researchers have explored the impact of specific working memory subsystems on children's mathematical performance, that is, the phonological loop (Swanson & Sachse-Lee, 2001; Wilson & Swanson, 2001) or the visual-spatial sketchpad (VSSP; Bull et al., 2008; Holmes, Adams, & Hamilton, 2008; McKenzie, Bull, & Gray, 2003; Rasmussen & Bisanz, 2005; Simmons, Singleton, & Horne, 2008). Some researchers have found that the VSSP is more important in mathematical performance for younger children, whereas the phonological loop is more important for older children (Holmes et al., 2008; Krajewski & Schneider, 2009; McKenzie et al., 2003). Thus, in younger children, measures of visual-spatial working memory may predict mathematical achievement independently of the linguistic or quantitative pathways.

> *Hypothesis 2: Linguistic, quantitative, and spatial attention pathways vary in their unique and relative contributions to later mathematical performance, depending on the demands of the tasks*

Because conventional tests of mathematical skills involve number system knowledge, the linguistic pathway is likely to be predictive of all mathematical outcomes. However, most conventional tests of mathematics presumably also require knowledge of numerical magnitude and skill at manipulating quantities. Furthermore, conventional tests typically engage attentional processes (DeStefano & LeFevre, 2004; Swanson, Jerman, & Zheng, 2008). The Pathways model predicts that to the extent that specific mathematical tasks involve different skills, the three pathways will vary in their unique and relative contribution to each task.

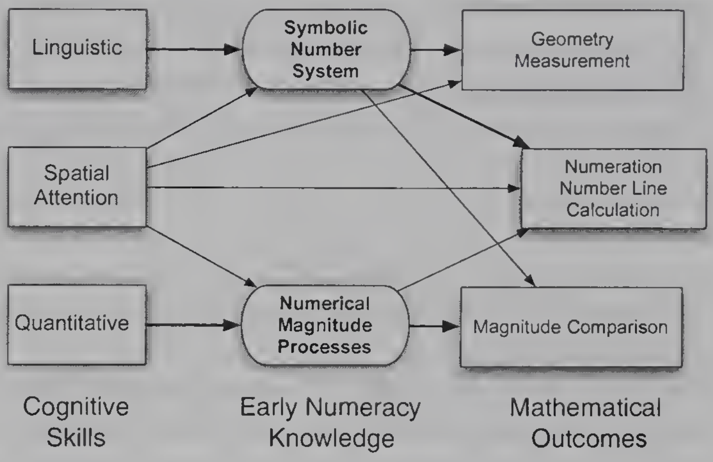


Figure 1. The Pathways model showing predicted relations among cognitive precursors, early numeracy skills, and mathematical outcome measures.

Figure 1 captures the hypotheses and predictions that arise from the Pathways model. Linguistic abilities form a skill pathway with children's symbolic number system knowledge. Quantitative abilities form a separate skill pathway linked with processing numerical magnitudes (i.e., quantity processing that does not require knowledge of the conventional number system). Spatial attention forms a third pathway. We assume that spatial attention will be important for the development of a variety of numerical and mathematical skills. As shown in the figure, we expected the three pathways to contribute differentially to the various outcomes that we assessed. Importantly, the figure reflects the general prediction from the model that the relative contribution of each pathway will depend on the task requirements. All outcomes should require some knowledge of the symbolic number system and hence the involvement of the linguistic pathway, but the quantitative pathway may not contribute uniquely for all outcomes.

Method

Data were drawn from a longitudinal project that spanned 4 years and included over 500 children and a wide variety of measures. In the present analyses, we focused on the 182 children who started the study in preschool (age 4) or kindergarten (age 5) and participated for 3 years. Several decisions guided choices of measures and analyses. First, we used standardized measures, based on age or grade norms as appropriate, whenever these were available. This decision allowed maximum comparability with other studies and provided a control for age-related effects. Second, for measures that did not have norms, we calculated z scores within age groups to use in correlational and regression analyses. Third, we focused on concurrent relations between (a) cognitive precursors and early numeracy measures from the 1st year of participation and (b) performance on conventional measures 2 years later.

Participants

A total of 196 preschool and kindergarten children were recruited from seven schools in three Canadian cities (Ottawa, Peterborough, and Winnipeg). Data were missing for 14 of these children on one or more measures in the 1st year of testing, and thus the present analyses are based on the 182 children (37 preschoolers and 145 kindergarteners) who had complete data for Year 1. The kindergarten children attended half-day classes in the schools. The preschool group attended half-day programs, either in the same schools (i.e., "junior" kindergarten in Ottawa and Peterborough) or in feeder preschools. Children in the preschool group (19 girls, 18 boys) ranged in age from 4:5 to 5:8 (years:months) at the time of the study, with a median age of 5:0. Children in the kindergarten group (70 girls, 75 boys) ranged in age from 5:4 to 6:6, with a median of 5:11 (including 1 boy who was 4:7 and had been allowed to start kindergarten 1 year early).

Two years later 122 of these children participated, 23 of the group who had started the study in preschool (11 girls, 12 boys) and 99 of the group who had started in kindergarten (44 girls, 55 boys). Attrition occurred because families moved ($n = 23$), consent was not obtained ($n = 24$), or the school declined to participate (one school, $n = 13$ children) and it was not possible to contact the families outside of the school. Thus, over 2 years the attrition rate because of all causes was 33%. A comparison of the children in the longitudinal sample and those who were not retained is presented as a preliminary analysis in the Results section.

Most of the children came from middle-class homes. The majority of families (87%) reported two adults living in the home. The majority of mothers had a university or postgraduate degree (56%) or had a diploma from a vocational college (18%). Only 2% had not finished high school. Most of the children spoke English at home (95%).

Materials and Procedures

Beginning in May of each year, children were tested by trained experimenters. Computer tasks and pencil-and-paper tasks were administered in two separate sessions lasting 15–30 min each or, in some cases, a single longer session. Tasks were performed in the same order in each session (paper or computer), but which session was conducted first varied because of the availability of experimenters and children.

Cognitive Precursors

Linguistic skills. Various aspects of linguistic development, such as phonological awareness, vocabulary, verbal reasoning, and listening comprehension (Aunola et al., 2004; Durand et al., 2005; Hecht, Torgesen, Wagner, & Rashotte, 2001; Krajewski & Schneider, 2009; Simmons et al., 2008), have been implicated as predictors of mathematical skill in the early school years. We hypothesized that receptive vocabulary would reflect children's ability to acquire vocabulary in the number system. Furthermore, acquiring the visual symbols (i.e., Arabic digits) and linking them to mental labels is similar to developing lexical or sublexical mappings when children are learning to read. Thus, we chose measures of vocabulary and phonological awareness to index linguistic skills.

Receptive vocabulary was measured with the Peabody Picture Vocabulary Test–Revised, Form B (Dunn & Dunn, 1997). This task is standardized by age with a mean of 100 and standard deviation of 15, and with a reported internal consistency (Cronbach's alpha) of .94. Phonological awareness was indexed by the elision subtest of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999). Children hear a word and are asked to say the word again, but omitting a sound (e.g., brat without the /r/). The test is age standardized with a mean of 10 and a standard deviation of 3, with Cronbach's alpha for 5-year-olds of .90. All children completed these measures when they were in kindergarten.

Quantitative knowledge. We assumed that quantitative knowledge reflects a child's ability to evaluate and discriminate quantities with minimal additional cognitive requirements (see also Halberda et al., 2008; Landerl, Bevan, & Butterworth, 2004; Willburger, Fussenegger, Moll, Wood, & Landerl, 2008). On this view, one way to assess quantitative knowledge is to measure children's ability to compare numerosities of sets of objects and determine which has more (e.g., Halberda et al., 2008). Another is to use symbolic representations of quantity, that is, Arabic digits or words, in number comparison tasks (e.g., Durand et al., 2005; Landerl et al., 2004) or to ask children to do magnitude comparisons of digits versus small quantities (Which is larger, 3 or **? e.g., Koontz & Berch, 1996). In both of these approaches, quantity knowledge is measured via a comparison. We chose a third approach to assess quantitative knowledge. When asked to indicate the quantity of small sets of objects, people are faster and more accurate on set sizes of 1–3 or 4 than on set sizes of 5 or larger (Mandler & Shebo, 1982). Moreover, response latencies are similar on small sets, and increase linearly on sets greater than 4. These data are consistent with the view that children and adults use a rapid pattern-matching process, *subitizing*, on small sets and counting-based processes on larger sets. Researchers have suggested that subitizing is a reasonable index of children's quantitative knowledge (Desoete & Grégoire, 2006; Landerl et al., 2004; van der Sluis, de Jong, & van der Leij, 2004; Willburger et al., 2008).

Children were shown arrays of one to six dots and were asked to state "how many" dots as quickly and accurately as possible. Children initiated the trial by pressing the space bar. The trial ended when the child stated the numerosity of the set and the experimenter pressed a key and then typed the response. There were three trials of each array size, each in a different random pattern. The median subitizing latency was computed from the latencies of the nine trials showing one, two, and three dots.

The children completed this task each year, but because subitizing was considered a precursor skill, only Year 1 data were used for model testing. However, cross-year consistency of subitizing speed was of interest for reliability and validity. The test–retest correlation in the larger longitudinal study was $r(237) = .46$, $p < .001$, between the 1st and 2nd years, and $r(228) = .30$, $p < .001$, between the 1st and 3rd years.

Spatial attention. Existing research on the role of working memory in mathematical development has indicated that the spatial span task (also known as Corsi or block span) is a useful index of visual-spatial working memory in young children (Bull et al., 2008; Holmes et al., 2008; McKenzie et al., 2003; Rasmussen & Bisanz, 2005; Simmons et al., 2008). In this task, the experimenter indicates a series of locations by touching each one briefly. The participant attempts to reproduce the ordered sequence. In contrast to verbal measures, such as digit span, research with adults has shown that verbal strategies are unhelpful on the spatial span task (Ridgeway, 2006). However, spatial and executive working memory resources are often difficult to distinguish empirically (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Thus, the spatial span task may index specific spatial working memory capacity, nonverbal strategies, and more general attentional capacity (Bisanz et al., 2005; Rasmussen & Bisanz, 2005). For present purposes, we refer to this complex of skills as *spatial attention*.

Children were shown a set of nine lily pads arranged in a pseudorandom pattern. For each trial, a frog appeared on a sequence of the lily pads (1 s per pad). Next, the child replicated the series by pointing to the lily pads. The experimenter clicked on each location that the child pointed to and the computer recorded the path. Following a demonstration trial, children attempted two trials at each series length, beginning with a series of two appearances, and progressing up to a series of six appearances. When a child responded incorrectly on both sequences of a given length, the task was terminated. The score consisted of the total number of correct sequences (maximum 12). Internal consistency, based on the two sequences at each series length, was .70 (Cronbach's alpha). The test–retest correlation between performance between the 1st and 2nd years (in the longitudinal study) was $r(182) = .44$, $p < .001$, and between the 1st and 3rd year was $r(228) = .56$, $p < .001$. Thus, children showed reasonable stability across years for this measure. For analyses of outcome measures, an average span was calculated using scores for all 3 years, or for the 1st and 3rd year for the 9 children who were missing a span score in Year 2. The internal reliability of this cross-year measure was .72 (Cronbach's alpha).

Early Numeracy Skills

We assumed that early numeracy skills (i.e., skills that children acquire before they start formal schooling) generally revolve around children's knowledge of numbers and quantities. We selected two tasks, number naming and nonlinguistic arithmetic, which capture important aspects of early numeracy and that we assumed would be maximally different in terms of requiring the quantitative versus the linguistic pathways.

Number naming (i.e., saying "twelve" when presented with "12") was assumed to reflect children's familiarity with Arabic number symbols. This task requires children to map Arabic digits to a phonological representation but does not require them to access the quantity or magnitude associated with the symbol (although, with practice, quantity may be accessed automatically). In Dehaene et al.'s (2005) model, naming numbers invokes internal representations associated with a linguistic code.

Children named Arabic numbers presented on the computer screen one at a time, in the following order: 3, 1, 7, 2, 9, 12, 15, 20, 22, 37, 64, 77, 95, 101, 210, 688, 1,500, 4,302. The task was terminated after three consecutive errors. Errors included instances in which child named the digits in numbers greater than 9 (e.g., said "one-five" for the number 15), said "I don't know," or produced only part of the number correctly (e.g., said "thirty-five" for the number 37). The score consisted of the number of correct responses (maximum 18).

In the nonlinguistic arithmetic task (Huttenlocher, Jordan, & Levine, 1994; Levine, Jordan, & Huttenlocher, 1992), children are shown an initial quantity that is then hidden. They watch as a transformation is applied to the hidden objects (i.e., some are visibly removed or added) and then reproduce the hidden number with their own objects. This task requires children to represent and mentally manipulate quantities, but does not require them to label those quantities with number names or to link the quantities to the Arabic symbols.

Children began with three practice trials in which the experimenter lined up a set of animals in front of a barn and then pushed them into the barn so that they were hidden from the child. The child was asked to model the number of hidden animals by using their set of animals. On the first three trials, the child simply matched the number of animals in the barn (3, 1, 5). On the next three trials, children were given addition problems: After the original set was moved to the barn, more animals were added (1 + 2, 2 + 3, 2 + 2). On the last three trials, children were given subtraction problems: After the original set was moved to the barn, some animals were moved out (2 − 1, 4 − 2, 5 − 4). The problems were presented in the same order for each child. The maximum possible score was 9, 3 for each type of problem. Percentages of correct trials were used in the analyses.

Outcome Measures

Children completed the outcome measures in the 3rd year of the project.

Conventional tests of mathematical knowledge. These tests included the Numeration, Geometry, and Measurement subtests from the KeyMath Test–Revised (Connolly, 2000) and the Calculation test from the Woodcock–Johnson Tests of Achievement–Revised (WJ–R; Woodcock & Johnson, 1989). These tasks were expected to capture many of the mathematical skills that children are assumed to learn in the first 2 years of school. Furthermore, these tasks were chosen to represent a range of measures that, although all central to mathematics, might vary in the extent to which they required linguistic versus quantitative knowledge.

The KeyMath Test has Canadian norms and the subtests are consistent with curricular guidelines. The Numeration subtest assesses quantity, order, and place value. Most of the items in the range for these children require knowledge of the symbolic number system. The Measurement subtest includes a variety of questions for which children are asked to compare quantities (e.g., size, extent, volume, mass) and choose the larger or smaller. The Geometry subtest includes some vocabulary items (e.g., sphere), some questions that require language-mediated processing of spatial arrays (e.g., point to the fourth-last item in the third row of the table), and some sequencing and patterning questions. For these three tests, grade-scaled normative scores were used ($M = 10$, $SD = 3$).

The Woodcock Calculation test is a traditional paper-and-pencil measure of factual and procedural computation skill (Woodcock & Johnson, 1989). Testing stopped after six incorrect or "I don't know" answers. This test has a median reliability of .85 and a 1-year test–retest correlation of .89 for Grades 2 through 4.

Research-based measures of mathematical knowledge. Children completed two additional mathematical measures, administered on laptop computers. These measures are commonly used in the research literature on children's mathematical development. In contrast to the conventional tests, intended to tap a range of knowledge, these research-based measures tap specific knowledge in greater detail.

The Number Line task requires that children select the appropriate position on a number line for numbers between 0 and 1,000. This number line measure is intended to capture both number system knowledge and children's internal quantity representations (Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003; cf. Moeller, Pixner, Kaufmann, & Nuerk, 2009). We hypothesized that this task relies heavily on the linguistic pathway because it requires understanding of symbolic numbers but may also index the extent to which children have coordinated symbolic and numerical magnitude knowledge. Children completed this task when they were in Grade 2.

Children initiated each trial by pressing a Go button, which displayed the target number and cursor in the image of a car. They used the mouse to move the cursor to the position on the displayed number line. Endpoints of the line were labeled 1 and 1,000. The computer recorded the location and the solution latency. Stimuli were chosen based on Laski and Siegler (2007) and were balanced with four targets between 0 and 100, four between 900 and 1,000, two targets from each other decade and distances matched from the endpoints. The targets [and matching items] were: 6 [994], 18 [982], 59 [991], 97 [903], 124 [876], 165 [835], 211 [789], 239 [761], 344 [656], 383 [617], 420 [580], 458 [542], and 500. Order of the 25 trials was randomized separately for each child.

Regression was used to calculate the relation between the actual values of the presented numbers and the locations that the child chose. Perfect performance would occur if each presented number were placed at the appropriate location (e.g., 500 at the center of the line). Various indices of performance can be calculated. In previous work, Siegler and colleagues have used mean r^2 values from linear regressions (regressing line location on actual number) as the index of fit. In our data, examination of the distribution of r^2 values showed a significant deviation from normality, with a value for skew of −.98 ($SE = .25$; 95% CI = .48). An alternative index is the slope of the regression line. Perfect placement of each number on the line would result in a slope of 1.0. Random placement would result in slopes close to 0. In the present research, slope values were distributed normally, with a value of skew of −.21 ($SE = .25$, 95% CI = .49). To allow comparison with previous work, analyses of both r^2 and slope values are presented.

Children also completed a symbolic magnitude comparison task (Landerl et al., 2004). Symbolic magnitude comparison is assumed to require both linguistic and quantitative knowledge (Holloway & Ansari, 2009). Children viewed pairs of single-digit numbers that varied in size and indicated which number was "more" by pressing an appropriate key. Of the 40 trials, on 20 the numerically larger number was also physically larger (i.e., congruent trials) and on the other 20, physical size and magnitude were incongruent. The split between the numbers was 1 for 24 trials (e.g., 2 vs. 3) and 5 for 16 trials (e.g., 2 vs. 7). Two practice trials preceded the experimental trials. Children viewed one of two randomized lists.

Children pressed the space bar to start the trial. After a 500-ms delay, the two numbers appeared simultaneously. The trial terminated after the child responds or if 3 s elapsed with no response. Response times and accuracy were recorded. The percentage of correct trials was used as the outcome measure.

Word reading. Children completed the Word Identification subtest of the Woodcock Reading Mastery Tests–Revised/Normative Update, Form G (Woodcock, 1998). Word Reading was used to examine the possibility that all complex cognitive tasks, not just mathematical ones, would involve all of the linguistic, spatial attention, and quantitative pathways. In accord with the predictions of the pathways model, we hypothesized that only the linguistic pathway, and not the quantitative pathway, would be related to word reading performance. In this test, children read isolated words. Split-half reliability coefficients for Grade 1 children are reported as .98. Scores are standardized by grade to have a mean of 100 and a standard deviation of 15. All children completed this test in Grade 1.

Results

Means, standard deviations, and correlations among tests are shown in Table 1 for the cognitive precursors and early numeracy skills, measured in Year 1, and in Table 2 for the outcome measures, measured in Year 3. Gender was correlated only with number naming and with the Number Line task. Boys named more numbers correctly than did girls, with mean z scores of .18 and −.31, respectively, $t(180) = -3.13$, $p < .001$. On the Number Line task, girls had smaller mean r^2 values than boys (.62 vs. .76), $t(93) = -2.77$, and flatter slopes (.48 vs. .61), $t(93) = 2.98$, $ps < .01$. Given these differences, gender was controlled in all regressions.

The expectation that children would subitize quantities up to 3, and count for 4, 5, or 6 was supported by accuracy and latency data. Both preschoolers and kindergarten children were virtually perfect on arrays of 1, 2, or 3 (98.3% and 99.5% correct, respectively). However, on arrays of 4–6, both groups made some errors, and younger children made more than older children (79% vs. 91%), $t(57.0) = -3.03$, $p < .01$. Median latencies were analyzed in a 2 (grade: preschool, kindergarten) × 6 (array size: 1–6) mixed analysis of variance. Data are shown in Figure 2. Younger children responded more slowly than older children (3,176 vs. 2,557 ms), $F(1, 172) = 20.68$, $MSE = 2{,}969{,}000$, $\eta_p^2 = .107$. Latencies increased with array size, $F(5, 860) = 300.77$, $MSE = 797{,}142$, $\eta_p^2 = .636$, and grade interacted with array size, $F(5, 860) = 5.39$, $\eta_p^2 = .03$. Children's latencies on arrays of one and

Table 1

*Correlations Among Variables in Year 1 (*M *and* SD *for the Standardized Tests) and Semipartial Correlation Coefficients From Multiple Regression Analyses of the Early Numeracy Measures (*N *= 182)*

	Correlations						Regressions	
	Precursors				Early numeracy			
	Vocabulary	Elision	Subitizing latency	Spatial span	Number naming	Nonlinguistic arithmetic	Number naming	Nonlinguistic arithmetic
Gender	.06	−.02	−.11	.00	.24**	.01	.23**	−.04
Vocabulary		.35**	−.09	.22**	.32**	.19*	.13*	.08
Elision			−.22**	.26**	.44**	.22**	.30**	.03
Subitizing latency				−.15	−.17*	−.38**	−.01	−.32**
Spatial span					.31**	.26**	.16*	.15*
Number naming						.21**	—	.05
Nonlinguistic arithmetic							.05	—
R^2							.32**	.21**
M	110.7	10.9	1,479	3.8	10.4	7.2		
SD	12.1	2.7	388	1.9	4.2	1.8		

Note. *M* and *SD* are in standardized units for vocabulary and elision, in ms for subitizing latency, and in number correct for the other measures. The *z* scores by grade were used in the correlational analyses.

$^{*}p < .05$. $^{**}p < .01$.

Table 2
Correlations Among Precursor Pathways and Outcome Tasks (M and SD)

	Word reading	Numeration	Calculation	Geometry	Measurement	Number line slope	Number line r^2	Magnitude Comparison
Gender	.05	.10	.01	−.03	−.13	.30**	.28**	.13
Spatial attention	.01	.41**	.48**	.44**	.44**	.45**	.44**	.32**
Linguistic	.44**	.62**	.51**	.52**	.43**	.44**	.57**	.32**
Quantitative	−.07	.26**	.31**	.22**	.26**	.31**	.37**	.39**
Word reading		.20*	.22*	.17	.01	.23*	.32**	.05
Numeration			.61**	.65**	.57**	.52**	.60**	.42**
Calculation				.39**	.48**	.45**	.56**	.47**
Geometry					.45**	.34**	.42**	.28**
Measurement						.34**	.37**	.35**
Number line slope							.89**	.37**
Number line r^2								.39**
N	109	120	118	119	119	94	95	116
M	118.0	12.1	100.7	11.2	10.9	.56	.70	81.3
SD	15.6	3.6	15.4	2.9	2.6	.21	.26	18.0

Note. *M* and *SD* are in standardized units for the tests, in slope units and proportion of variance for the Number Line task, and in percentage correct for the magnitude comparison task.
*$p < .05$. **$p < .01$.

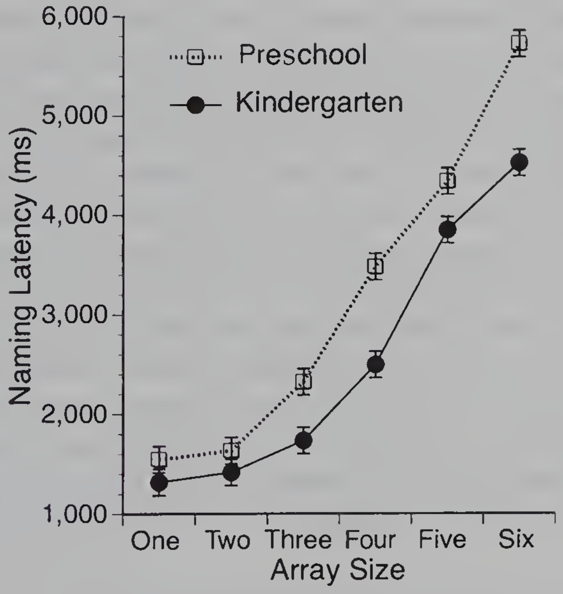


Figure 2. Means of median-naming latencies by array size.
Note. Whiskers are the 95% confidence intervals based on the mean square error for the interaction of array size and grade.

two dots were not different at either age level. However, three dots were named more slowly than two dots for both the preschool and the kindergarten children. Preschool children were slower on all array sizes than kindergarten children, but the differences were larger for arrays of 4, 5, and 6 than for 1, 2, and 3. Median latencies on arrays of 1–3 were used as the index of subitizing latency.

The performance of children who only participated in Year 1 and those who were retained throughout the study was compared for the six Year 1 measures. Only two differences were notable. First, among the preschool children, those who dropped out had somewhat slower subitizing latencies than those who continued (1.5 vs. 2.0 s), $t(14.8) = 2.12$, $p = .052$. Inspection of the data indicated that a few children in the group who were not retained had very slow latencies. Second, among kindergarten children, those who dropped out had lower scores on the nonlinguistic arithmetic task than those who continued in the study (6.8 vs. 7.8), $t(71.6) = 2.92$, $p < .01$. Inspection of the data suggests that there were relatively more children with perfect scores in the retained group than in the group who were not retained but that the lower range of performance was not restricted. These differences were unlikely to have biased the results in favor of our hypotheses but they might have resulted in an underestimate of the importance of the quantitative pathway.

Tests of the Pathways Model

Hypothesis 1: Linguistic, quantitative, and spatial attention pathways contribute independently to early numeracy

As shown in Table 1, all four precursors (vocabulary, elision, subitizing, and spatial span) were significantly correlated with both early numeracy measures, number naming, and nonlinguistic arithmetic. Consequently, multiple regressions were used to test the model, as shown in the last two columns of Table 1. These show strong support for the four predictions based on Hypothesis 1. First, linguistic skills (vocabulary, elision) predicted unique variability (as reflected by semipartial correlations) in number naming but not in nonlinguistic arithmetic. Second, quantitative skill (subitizing latency) predicted significant unique variability in nonlinguistic arithmetic but not in number naming. Third, although the simple correlation between number naming and nonlinguistic arithmetic was significant, neither accounted for unique variability in the corresponding regression. Fourth, spatial attention was a unique predictor of variability in both early numeracy measures. Thus, the patterns of performance support the view that linguistic and quantitative pathways are distinct predictors of early numeracy skills that assess number system knowledge and numerical magnitude processing, respectively. Spatial attention forms a third independent pathway that is related to both types of early numeracy performance.

We used factor analysis to compute separate linguistic and quantitative predictors for use in longitudinal analyses. The cognitive precursors (elision, vocabulary, subitizing speed) and the early numeracy measures (number naming and nonlinguistic arithmetic) were entered into a principal components analysis. Two factors emerged that accounted for 63% of the variance. The first factor included the linguistic measures, with factor loadings of .76, .74, and .76 for elision, vocabulary, and number naming, respectively. The second factor included subitizing latency and nonlinguistic arithmetic, with loadings of −.81 and .79. Factor scores were used in the longitudinal regressions to index the linguistic and quantitative pathways.

Hypothesis 2: Linguistic, quantitative, and spatial attention pathways vary in their unique and relative contributions to later mathematical performance, depending on the demands of the tasks

Table 3 contains results from multiple regressions in which gender and predictors for each of the three pathways are used to account for variability in performance on conventional and research-based mathematical outcomes, as well as word reading. Three observations are particularly noteworthy. First, the proportion of variability accounted for by these predictors is substantial, ranging from .26 to .56 for the mathematical outcomes. Second, the amount of variability predicted uniquely by each pathway varies, depending on the outcome. For example, all three pathways are implicated in performance on Numeration, Calculation, and the Number Line tasks, but the quantitative pathway is not involved in Geometry and Measurement and spatial attention is not involved in Magnitude Comparison. Perhaps not surprisingly, the linguistic pathway alone accounts for significant variability in Word Reading, but it also is the only variable to be implicated in all of the mathematical outcomes. Third, semipartial correlations are higher for the linguistic pathway than for the quantitative and spatial attention pathways for six of seven mathematical outcomes, indicating that the general sorts of language-related skills reflected in elision, vocabulary, and number naming are a key component in tests that typically are assumed to measure mathematical knowledge and skill. Thus, all three pathways are involved in perfor-

Table 3
Pathways for Differential Mathematical Outcomes as Indexed by Semipartial Correlation Coefficients From Multiple Regressions

		Mathematical outcomes						
		Conventional tests				Research-based measures		
Predictors	Word reading	Numeration	Calculation	Geometry	Measurement	Number line slope	Number line r^2	Magnitude comparison
Gender	.02	.08	.03	−.04	−.13	.26**	.21**	.10
Spatial attention	−.12	.14*	.27**	.23**	.24**	.29**	.24**	.12
Linguistic	.45**	.54**	.39**	.41**	.32**	.32**	.47**	.24**
Quantitative	−.01	.19**	.20**	.08	.12	.17*	.24**	.30**
R^2	.21**	.48**	.44**	.36**	.32**	.40**	.56**	.26**
N	109	119	117	118	118	93	93	116

Note. Predictor measures are provided in rows and criterion measures in columns.
$^*p < .05$. $^{**}p < .01$.

mance on some, but not all mathematical tasks, and the degree of involvement varied across tasks.

Discussion

We developed and tested a model of the relations among cognitive precursors, early numeracy knowledge, and the subsequent performance of conventional and research-based tests of mathematical skill. The first hypothesis was that linguistic, quantitative, and spatial attention would form separate skill pathways. In support of this hypothesis, we found that vocabulary and elision (linguistic skills) were related to number naming, whereas subitizing latency (quantitative skill) was related to nonlinguistic arithmetic. Linguistic skills did not predict nonlinguistic arithmetic and subitizing did not predict number naming, supporting the proposed distinction between these pathways. Accordingly, numerically relevant linguistic and quantitative skills formed two distinct factors for these young children.

Spatial attention showed significant relations to both number naming and processing of numerical magnitude, in support of a third pathway. The finding that spatial attention is related to the development of a variety of early mathematical skills is consistent with recent findings in which similar measures of nonverbal working memory were related to children's mathematical performance in the first few years of school (Bull et al., 2008; Holmes et al., 2008; Krajewski & Schneider, 2009; Simmons et al., 2008). More generally, attentional processes may be important in children's ability to manage the complex requirements of mathematical tasks (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007).

There was also support for the second hypothesis that the roles of the linguistic, quantitative, and spatial attention pathways would vary depending upon the mathematical outcomes. The spatial attention pathway accounted uniquely for moderate amounts of variability on six of seven mathematical tasks but was never the dominant predictor. The linguistic pathway accounted uniquely for variability in all of the mathematical outcomes but the strength of the relation varied such that it was weakest with magnitude comparison and strongest with numeration. Thus, as predicted, the relative importance of the pathways depends on the extent to which number system and numerical quantity knowledge is used in the mathematical task. For example, magnitude comparison requires only that children know the symbol–quantity correspondence for single-digit numerals from 1 to 9. In contrast, the numeration test requires knowledge of three-digit numbers, including advanced understanding of place value. Similarly, the linguistic pathway was a strong predictor of the linearity of children's number line knowledge for the numbers from 1 to 1,000. This task requires children make a connection between symbolic representations (e.g., 864) and a spatial extent. They would not be able to attempt the task without a reasonably good understanding of the Arabic symbols for numbers to 1,000.

For the quantitative pathway, the strength of the relation with the outcomes overall was modest. Of interest were the relative patterns for the linguistic and quantitative pathways. Here, two extremes emerged. For geometry and measurement, the linguistic pathway was uniquely predictive but the quantitative was not, even though the quantitative pathway was correlated with performance on these outcomes. In contrast, for magnitude comparison, the quantitative pathway was numerically but not statistically more strongly related to performance than the linguistic pathway. This result is consistent with some existing work suggesting that performance on the magnitude comparison task reflects children's skill in linking together symbolic and nonsymbolic numerical processes (DeSmedt et al., 2009; Holloway & Ansari, 2009; Mundy & Gilmore, 2009; Rousselle & Noël, 2007). The finding that linguistic, quantitative, and spatial attention pathways vary in their relation to mathematical outcomes has two notable implications. First, the Pathways model could prove helpful in understanding why some children might do well at one mathematical task but not another. Second, the model may prove useful for determining the psychological demands of tasks that are designed ostensibly to measure mathematical skill.

Connections to Research With Special Populations of Children

The Pathways model is consistent with research on special populations of children who have deficits related to mathematics. Using the model, many children can be grouped according to whether their deficits are in linguistic or quantitative processing. For example, children with specific language impairments (SLI) do poorly on mathematical tasks that have a linguistic component (and thus involve knowledge of the number system) but perform similarly to peers without SLI on tasks where quantitative knowledge is most relevant (Donlan, 2007). In contrast, children with Williams syndrome (who have both spatial difficulties and mathematical delays) appear to use their relatively stronger linguistic skills to do mathematical tasks and hence show specific patterns of weaknesses on tasks that require quantitative processing (Ansari et al., 2003). Similarly, children with spina bifida and cerebral palsy, who have motor deficits and compromised mathematical skills, appear to have deficits in the quantitative pathway (Barnes et al., 2005). Thus, the pathways model will be useful in understanding this dissociation in linguistic and quantitative performance for children with specific deficits in mathematics. We also speculate that children who have a specific mathematics disability (i.e., dyscalculia) might show particular weakness in the quantitative pathway, but overall average or better performance in the linguistic pathway. In contrast, children who show deficits in both mathematics and reading (i.e., are dyscalculic and dyslexic) are weak in both pathways (Jordan et al., 2007; Willburger et al., 2008). Finally, children with attention deficit hyperactivity disorder often have mathematical difficulties (Zentall, 2007). According to the pathways model, we might expect those children to have deficits in the spatial attention pathway, but show typical variability in relevant linguistic and quantitative skills.

Future Considerations

Further work is needed to refine and develop the Pathways model theoretically, empirically, and methodologically. In the Pathways model, it was reasonable to use spatial span to index attention because we linked our model to Dehaene's neuropsychological account of the parietal circuits that support numerical processing. However, working memory and attentional processing more generally involve the coordination of processes and knowledge that go beyond the parietal circuits. For example, central executive resources also involve the frontal lobe (Bull et al., 2008; Espy & Bull, 2005). Furthermore, working memory is assumed to involve coordination of information from the world and from memory. As the Pathways model is further developed to encompass learning trajectories and related processes, a more detailed account of how working memory and related attentional processes are important will become critical (Geary et al., 2007; Swanson et al., 2008).

The issue of how to operationalize the core quantity skill, as processing of small exact numerosities versus as approximate magnitudes (or both), also remains an important area for further research. Our decision to operationalize quantitative knowledge as subitizing latency (i.e., efficiency of identifying small exact quantities) was based on Butterworth's view that processing exact numerosities is the source of mathematical difficulties for children with dyscalculia (Butterworth, 1999, 2005). An alternative view is that mathematical difficulties reflect limitations in children's speeded access to the links between quantities and number names (see also Holloway & Ansari, 2009; Koontz & Berch, 1996; Rousselle & Noël, 2007). Thus, it may be the ability to create a mapping between symbolic knowledge and numerical quantity representations that is crucial in mathematical learning (Ansari, 2008). Furthermore, subitizing has not been universally accepted as a core quantity skill. An alternative view is that subitizing is a visual–attentional process that is not specific to numerical tasks (see also Trick & Pylyshyn, 1994). For example, children with cerebral palsy, especially those with right-hemisphere damage, are impaired on both subitizing and counting relative to normally developing children and have impaired pattern recognition processes that are not numerically specific (Arp & Fagard, 2005). Arp and Fagard (2005) concluded that counting and subitizing both reflect global, rather than numerically specific, processes. In this view, subitizing processes may be recruited to support quantitative tasks and thus become more closely related to quantitative processing over time. Clearly, more evidence is needed to determine the relations among subitizing, nonsymbolic magnitude comparisons, and the development of various mathematical abilities and knowledge (Ansari, 2008; Holloway & Ansari, 2009; Rousselle & Noël, 2007).

As formulated, the Pathways model does not address the question of how knowledge from the linguistic and quantitative pathways is integrated, although working memory may play an important role in this process (Krajewski & Schneider, 2009; Swanson et al., 2008). Children's knowledge of mathematics, at least for normally developing individuals, presumably reflects the interplay of linguistic and quantitative skills because both are recruited by most complex mathematical tasks.

Conclusion

We proposed and tested a model of how cognitive skills are implicated in the development of children's mathematical knowledge. The Pathways model includes three forms of knowledge: quantitative, linguistic, and spatial attention, and it specifies the initially distinct role each of these plays in early numeracy acquisition. We showed that the three pathways relate differentially to conventional tests of mathematical knowledge, in support of the view that mathematical development is componential. The model connects research from adults, neuroimaging, clinical populations, and normal development. Furthermore, across various mathematical outcomes (e.g., geometry, measurement, calculation, numeration), the Pathways model highlights the need to understand the fundamental underlying skills that contribute to all of these specific domains.

References

Ansari, D. (2008). Effect of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience, 9*, 278–291.

Ansari, D., Donlan, C., Thomas, M. S. C., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology, 85*, 50–62.

Arp, S., & Fagard, J. (2005). What impairs subitizing in cerebral palsied children? *Developmental Psychobiology, 47*, 89–102.

Aunio, P., Niemivirtaa, M., Hautamaki, J., Van Luit, J. E. H., Shi, J., & Zhang, M. (2006). Young children's number sense in China and Finland. *Scandinavian Journal of Educational Research, 50*, 483–502.

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to Grade 2. *Journal of Educational Psychology, 96*, 699–713.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.

Barnes, M. A., Smith-Chant, B. L., & Landry, S. H. (2005). Number processing in neurodevelopmental disorders: Spina bifida myelomenigocele. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 299–313). New York: Psychology Press.

Berch, D., & Mazzocco, M. (2007). *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities*. Baltimore: Paul H. Brookes.

Bisanz, J., Sherman, J. L., Rasmussen, C., & Ho, E. (2005). Development of arithmetic skills and knowledge in preschool children. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 143–162). New York: Psychology Press.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205–228.

Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. New York: Free Press.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*, 3–18.

Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 103*, 4693–4698.

Connolly, A. J. (2000). *KeyMath–Revised/Updated Canadian Norms*. Richmond Hill, Canada: PsyCan.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14*, 218–224.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2005). Three parietal circuits for number processing. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 433–453). New York: Psychology Press.

DeSmedt, B., Swillen, A., Devriendt, K., Fryns, J. P., Verschaffel, L., & Ghesquière, P. (2007). Mathematical disabilities in children with velo-cardio-facial syndrome. *Neuropsychologia, 45*, 885–895.

DeSmedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematical achievement. *Journal of Experimental Child Psychology, 103*, 469–479.

Desoete, A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences, 16*, 351–367.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386.

Donlan, C. (2007). Mathematical development in children with specific language impairments. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 151–172). Baltimore: Paul H. Brookes.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III*. Circle Pines, MN: American Guidance Service.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91*, 113–136.

Espy, K. A., & Bull, R. (2005). Inhibitory processes in young children and individual variation in short-term memory. *Developmental Neuropsychology, 28*, 669–688.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying deficits in children with mathematical learning disabilities. *Child Development, 78*, 1343–1359.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9*, 6–10.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computational skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79*, 192–227.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17–29.

Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology, 20*, 272–289.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123*, 284–296.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Disabilities Research & Practice, 22*, 36–46.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*, 153–175.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition, 2*, 1–24.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516–531.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8- and 9-year-old students. *Cognition, 93*, 99–125.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison *Child Development, 78*, 1723–1743.

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53*, 72–103.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*, 451–459.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 11*, 1–22.

Mazzocco, M. M. M., & McCloskey, M. (2005). Math performance in girls with Turner or fragile X syndrome. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 269–297). New York: Psychology Press.

McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*, 93–108.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis *Journal of Experimental Psychology: General, 130*, 621–640.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*, 503–515.

Mundy, E., & Gilmore, C. (2009). Children's mappings between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*, 490–502.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*, 165–184.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157.

Ridgeway, D. (2006). Strategic grouping in the spatial span memory task. *Memory, 14*, 990–1000.

Rousselle, L., & Noël, M. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude. *Cognition, 102*, 361–395.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–243.

Simmons, F., Singleton, C., & Horne, J. (2008). Brief report–phonological awareness and visual-spatial sketchpad functioning predict early arithmetic attainment: Evidence from a longitudinal study. *European Journal of Cognitive Psychology, 20*, 711–722.

Spelke, E., & Dehaene, S. (1999). Biological foundations of numerical thinking: Response to T. J. Simon (1999). *Trends in Cognitive Sciences, 3*, 365.

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*, 343–379.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294–321.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision *Psychological Review, 101*, 80–102.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology, 87*, 239–266.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *The Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Willburger, E., Fussenegger, B., Moll, K., Wood, G., & Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learning and Individual Differences, 18*, 224–236.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities, 34*, 237–248.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests–Revised*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised*. Allen, TX: DLM Teaching Resources.

Zentall, S. S. (2007). Math performance of students with ADHD: Cognitive and behavioral contributors and interventions. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 219–243). Baltimore: Paul H. Brookes.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1768–1786

How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition

Clarissa A. Thompson
The University of Oklahoma

John E. Opfer
The Ohio State University

How does understanding the decimal system change with age and experience? Second, third, sixth graders, and adults (Experiment 1: N = 96, mean ages = 7.9, 9.23, 12.06, and 19.96 years, respectively) made number line estimates across 3 scales (0–1,000, 0–10,000, and 0–100,000). Generation of linear estimates increased with age but decreased with numerical scale. Therefore, the authors hypothesized highlighting commonalities between small and large scales (15:100::1500:10000) might prompt children to generalize their linear representations to ever-larger scales. Experiment 2 assigned second graders (N = 46, mean age = 7.78 years) to experimental groups differing in how commonalities of small and large numerical scales were highlighted. Only children experiencing progressive alignment of small and large scales successfully produced linear estimates on increasingly larger scales, suggesting analogies between numeric scales elicit broad generalization of linear representations.

The ratio structure of the decimal system—where "1" denotes a quantity 1/10 of 10, "10" a quantity 1/10 of 100, "100" a quantity 1/10 of 1,000, and so forth—may apply to an infinity of numbers, but even over a lifetime, experience of symbolic numbers is finite. And more, experiences are also systematically biased, with numerals in the first order of magnitude (1–9) appearing much more frequently than numerals in the second order (10–99), which occurs more frequently than in the third order (100–999), and so forth (Dehaene & Mehler, 1992). This cross-linguistic regularity in frequency of symbolic numbers—observed in languages as diverse as American English, Catalan, Dutch, French, Japanese, Kannada, and Spanish—has been observed across forms of notation (Arabic or written number words), for cardinal numbers (1, 2, 3) as well as ordinals (first, second, third), and in both text and speech.

This cross-linguistic regularity in frequency of symbolic numbers has two interesting implications for how children might develop their representations of numerical magnitude. The first developmental implication is that relative frequencies of numerals in the environment place a constraint on how broadly children extend their understanding of the decimal system. Specifically, children learning to map numerals to magnitudes would be expected to fail to realize that ratios that hold at smaller, more familiar orders of magnitude (e.g., relative magnitudes of 15 cherries and 100 cherries) extend to larger orders of magnitude (e.g., relative magnitudes of 1,500 and 10,000) with which they are less familiar.

The second implication is that if developing representations of large numerical magnitudes are constrained by poverty of input described by Dehaene and Mehler (1992), no inherent cognitive constraint would prevent rapid and broad representational changes from occurring in children. The crucial source of change in this account is children encountering information that relations among numbers at large orders of magnitude are similar to relations at small orders of magnitude (e.g., 150:1,000::15:100), much like relations at small orders of magnitude are similar to each other regardless of units involved (e.g., 15 cherries:100 cherries::15 pears:100 pears). If true, such findings

---

The authors would like to thank the administration, teachers, students, and parents from the Worthington School District in Columbus, OH, as well as the Bentworth School District in Bentleyville, PA. Further, the authors would like to thank Julia Kennedy for her help in data collection, Chris Young for his contributions to the "change point" analyses in the General Discussion section, and anonymous referees who provided helpful comments on earlier versions of the manuscript. Portions of these data were presented at the 2007 meeting of the Cognitive Development Society as well as the 30th annual conference of the Cognitive Science Society.

Correspondence concerning this article should be addressed to Clarissa A. Thompson, Department of Psychology, The University of Oklahoma, 455 W. Lindsey Street, Dale Hall Tower, Room 727, Norman, OK 73019. Electronic mail may be sent to cat3@ou.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0012

would be scientifically important because it could reconcile two sets of seemingly contradictory findings—slow rate of representational changes observed in cross-sectional studies (e.g., Siegler & Opfer, 2003; logarithmic-to-linear switch in numeric representations between second and fourth grade on 0–1,000 number line problems) and one-trial representational changes observed in microgenetic studies (e.g., Opfer & Siegler, 2007; second graders' adoption of a linear representation after maximally discrepant feedback is presented). That is, representational change in the domain of numbers may be driven by children's ability to draw an analogy between small and large orders of magnitude, and previous research has indicated that direct feedback may highlight this analogy of the decimal system for children (Opfer & Siegler, 2007).

Previous evidence supporting the first conjecture about children's developing understanding of this property of the decimal system—as well as potential for improvement coming from comparison of small and large orders of magnitude—comes from interindividual variability in magnitudes children associate with a given number. When asked to place numbers on a number line flanked by 0 and 100, for example, second graders provided Siegler and Opfer (2003) with a set of estimates that were highly accurate and increased linearly with numeric value. When asked to place the same numbers on a number line flanked by 0 and 1,000, however, these same children provided estimates that were much less accurate, with estimates increasing logarithmically with numeric value. Thus, across contexts, the magnitude "15" was estimated as having different relations to 100 and 1,000. Moreover, this variability appeared to reflect variability in children's thinking about numeric magnitude rather than "noise" attributable to measurement error: Given no feedback, children's numerical magnitude estimates on a number line are highly stable from trial to trial (Opfer & Siegler, 2007; Opfer & Thompson, 2008; Thompson & Opfer, 2008).

To address these two implications more directly, Experiment 1 was designed to test how broadly children spontaneously apply linear representations on number lines with large numeric anchors (e.g., 0–1,000, 0–10,000, and 0–100,000). To test whether analogy could serve as a mechanism of developmental change across numeric representations, Experiment 2 examined whether progressive alignment—a means of fostering analogies in young children (Gentner, Loewenstein, & Hung, 2007; Kotovsky & Gentner, 1996)—led numeric representations used at small numerical scales (0–100) to be generalized to progressively larger numerical scales (0–1,000, 0–10,000, and 0–100,000) simply by highlighting perceptual similarities across these numeric contexts.

In the next sections, we will detail (a) development of numeric representations across the life span, (b) analogy as a mechanism of developmental change across numeric representations, and (c) more specific empirical questions that were examined in the present studies.

Development of Numerical Representations

Children normally improve their understanding about magnitudes denoted by symbolic numerals across a wide range of tasks (e.g., number line estimation, number categorization, magnitude judgments, etc.). Development of numerical representations appears to occur iteratively, with parallel developmental changes occurring over many years and across many contexts (Siegler, Thompson, & Opfer, 2009). For example, sixth graders' estimates on 0–1,000 number lines increase linearly (Siegler & Opfer, 2003), whereas second graders' estimates increase logarithmically (Opfer & Siegler, 2007; Opfer & Thompson, 2008; Siegler & Opfer, 2003; Thompson & Opfer, 2008). On 0–100 number lines, second graders' estimates increase linearly (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003), whereas kindergartners' estimates increase logarithmically (Siegler & Booth, 2004). Asked to estimate 0–10 chips from a pile of chips, kindergartners' estimates increase linearly, but preschoolers' estimates increase logarithmically (Opfer, Thompson, & Furlong, 2010). Further evidence for logarithmic-to-linear shifts in numerical magnitude representations have also been discovered when (a) children were asked to estimate candies in a container, (b) estimate salaries given in fractional notation, (c) provide answers to arithmetic problems, (d) make measurements of novel units, and (e) categorize and compare symbolic numbers (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & DeVries, 2008; Opfer & Thompson, 2008; Thompson & Opfer, 2008).

Children's initial expectations that numerical magnitudes increase logarithmically are theoretically interesting because these expectations are consistent with Fechner's law. That is, just as sensations increase logarithmically with stimulus intensity (Fechner's law), representations of numeric magnitude also increase logarithmically with actual value (Dehaene, 2007). For instance, the difference between 1 and 10 cherries seems larger (or is more quickly detected) than the difference between 101 and 110 cherries, much as these numbers would be spaced on a logarithmic ruler (Dehaene, Dehaene-Lambert, & Cohen, 1998). This logarithmic representation of numerical magnitude is not unique to young children making estimates on number lines but is widespread across tasks (estimation of set sizes, numeric comparisons, random number generation, and economic games), across species (pigeons, rats, nonhuman and human primates), and across age groups (infants, children, and time-pressured adults; Banks & Hill, 1974; Brannon, 2005; Feigenson, Dehaene, & Spelke, 2004; Furlong & Opfer, 2009; Gallistel & Gelman, 1992; Moyer & Landauer, 1967; Roberts, 2005; Xu & Spelke, 2000).

Children's learned expectations that numerical magnitudes increase linearly are not just theoretically interesting but also have practical effects on their school achievement. For instance, linearity of children's number line estimates correlates strongly with how quickly they compare magnitudes (e.g., is 5 or 7 greater; Laski & Siegler, 2007), their ability to learn solutions to unfamiliar addition problems (Booth & Siegler, 2008), and their overall scores on mathematics achievement tests (Booth & Siegler, 2006; Siegler & Booth, 2004). Playing numeric board games (akin to Chutes and Ladders) has also been shown to impact preschoolers' ability to compare numerical value and to produce accurate answers to arithmetic problems (Griffin, Case, & Siegler, 1994; Ramani & Siegler, 2008; Siegler & Ramani, 2008, 2009). Thus, not only does the development of linear representations of numerical magnitude present an interesting scientific problem, it is also one with important educational consequences, an importance underlined by recommendation of the National Mathematics Advisory Panel to embrace practices that foster ''number sense,'' including ability ''to estimate orders of magnitude'' (U.S. Department of Education, 2008, p. 18).

Analogy as a Mechanism of Developmental Change Across Numeric Representations

By what mechanism might children abandon use of a logarithmic representation of numerical magnitude for use of a linear representation? Across many areas of cognitive development, an important mechanism for representational change is analogy, where elements in a target domain may share structural relations with a base domain despite lack of perceptual overlap among elements in the target and base domains (Chen & Klahr, 1999; Gentner et al., 1997; Holyoak & Thagard, 1995; Opfer & Siegler, 2004, 2007).

Although young children may initially have difficulty solving analogical reasoning problems (see Abdellatif, Cummings, & Maddux, 2008, for a recent review of the factors affecting the development of analogical reasoning in young children), a number of factors can improve analogical reasoning in children such as: (a) asking children to solve problems with familiar examples or familiar relations (e.g., loaf of bread:single slice of bread::lemon:___; Goswami & Brown, 1990), (b) providing direct instruction on higher order relations through the use of relational language (e.g., ''Daddy, Mommy, and Baby'' for sizes big, medium, and small; Rattermann & Gentner, 1998), (c) explaining classic analogies of the format A:B::C:D (e.g., noting the relation between A and B and the similarity between A and C and B and D; Alexander, Willson, White, & Fuqua, 1987), (d) training children to solve a base problem and then asking them to transfer this knowledge to solve a related target problem (e.g., children are told a story about a genie who moves jewels from one bottle to another by rolling a ''magic carpet'' into a tube shape, and then the children are asked to solve the problem of moving gumballs from one bowl to another; Holyoak, Junn, & Billman, 1984), and (e) comparing across various instances (e.g., Gentner & Namy, 1999).

Structure-mapping theory (Gentner, 1983) guides our understanding of children's analogical reasoning ability in the current experiments. According to structure-mapping theory, children form analogies by aligning representational elements between a base and target domain. This alignment process facilitates transfer of information from base to target through children's comparison of surface-level features. This comparison process leads to subsequent highlighting of common underlying relational structure shared by base and target (Kurtz, Miao, & Gentner, 2001).

In progressive alignment (Gentner et al., 2007; Kotovsky & Gentner, 1996)—a specific instantiation of structure-mapping theory principles—comparisons made between highly similar elements can promote subsequent analogical matches of lower overall similarity. Analogical matches are thus promoted because the process of aligning surface-level elements highlights common underlying relational structure. The alignment process makes this structure more salient and easily mapped from base to target domain even when participants are presented with less surface-similar matches.

According to Kotovsky and Gentner (1996), progressive alignment acts as a mechanism of representational change by allowing children to make similarity comparisons over concrete, perceptual similarities (e.g., monotonic increase in size across differently shaped stimuli). Then, these similarity comparisons facilitate children's ability to notice higher order relational commonalities across stimuli possessing fewer surface-level features in common (e.g., increase in size as compared to saturation of color across differently shaped stimuli). Thus, progressive alignment allows children to recognize "richer and deeper" abstract relational similarity uniting their mental representations that may not have been immediately apparent before similarity comparisons were made (Kotovsky & Gentner, 1996).

Our general perspective on representational change, drawn from computational models of cognition (Doumas, Hummel, & Sandhofer, 2008; Gentner, 1983; Hummel & Holyoak, 2003), historical changes in scientific concepts (Gentner et al., 1997; Holyoak & Thagard, 1995), and microgenetic studies of children's concept learning (Opfer & Siegler, 2004, 2007) immediately suggested analogy as a candidate mechanism for development of numerical representations. This hypothesis seemed especially likely to be true for number line estimates, where structure of elements in a number line problem (e.g., relation between right anchor and number to be estimated) offers a systematic regularity over orders of magnitude (e.g., 150:1,000::15:100).

Analogy was also suggested as a candidate mechanism of representational change for second-grade students in a recent number line training study conducted by Opfer and Siegler (2007). Although these second-grade students produced a logarithmic series of estimates on 0–1,000 number lines during pretest, when children received feedback on the placement of 150 (maximally discrepant point between logarithmic and linear functions forced to pass through 0 and 1,000), children immediately generated a linear series of estimates for all other numbers in the 0–1,000 numeric scale that were tested. This abrupt and broad generalization of learning—characteristic of analogical mapping more broadly (Chen & Klahr, 1999; Gentner, Holyoak, & Kokinov, 2001; Holyoak & Thagard, 1995; Opfer & Siegler, 2004)—led Opfer and Siegler to speculate that children had improved their performance by mapping commonalities between the less familiar, larger scale (0–1,000) to the more familiar, smaller scale (0–100) where children already possessed a linear representation.

In this article, we provided a novel test of Opfer and Siegler's (2007) hypothesis that analogy provides a mechanism for developmental changes in representations of numerical magnitude. Our test was unique in two important respects. First, if children can draw an analogy between 0–100 and 0–1,000 number line problems, feedback that Opfer and Siegler provided in the 0–1,000 scale would not be necessary for children to improve their performance on those problems. To test this issue, our strategy for eliciting analogies from children was to manipulate factors that would lead to comparisons of number line problems, and we otherwise avoided giving children feedback on target problems. Second, if children can extract general structural relations hypothesized by Opfer and Siegler (i.e., 150:1,000::15:100), they should generalize far outside the training space (e.g., from a 0–100 number line to a 0–100,000 number line), an issue that Opfer and Siegler did not address.

Further, we have proposed that possession of a numeric analogy between familiar, small numeric scales (0–100) and unfamiliar, larger numeric scales (0–1,000 and larger) can prompt children to scale up a linear representation of number to unfamiliar scales. Thus, familiarity with numeric context (0–100) plays an important role in our explanation of analogy as a mechanism of representational change, but another model of children's estimation performance, the segmented linear model (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008), suggests that a very different type of familiarity affects children's estimates.

According to Ebersbach et al.'s (2008) segmented linear model, children's familiarity with numbers (indexed by their counting ability within a particular numeric scale) is a good predictor of shape of the participants' mental number line (e.g., linear or logarithmic). That is, if participants are familiar with the numeric context in which they are asked to make estimates, the best fitting function of participants' number line estimates should be linear with a greater slope than the scale with which they are unfamiliar, thereby leading to a pattern of estimates that appears logarithmic. Within this account, apparent fit of the logarithmic model is illusory, whereas a segmented linear model, consisting of two simple linear regression functions that intersect at a "change point" provides a better model fit. Most impressively, "change point" in this segmented linear model was correlated with children's counting range, suggesting the apparent fit of the logarithmic function may be a by-product of children's familiarity. In this article, we investigate whether the assumptions of the segmented linear model hold in larger numeric scales (e.g., 0–10,000 and 0–100,000).

Issues Examined in Present Studies

The central purpose of the present studies was to examine how broadly children spontaneously apply linear representations on number lines with large numeric anchors (e.g., 1,000, 10,000, and 100,000; Experiment 1) and to determine whether aligning structure of small and large numeric scales will prompt children to generalize linear representations more broadly than they would without such alignment (Experiment 2).

The purpose of Experiment 1 was to determine age differences in underlying representations that children and adults used when making estimates on number lines of increasing numerical magnitude and track the developmental trajectory of participants' linearization of estimates on number lines of increasing magnitudes. To investigate these age differences in numerical representations, we examined second-, third-, and sixth-grade students' number line estimation performance in large numerical contexts (e.g., 0–1,000, 0–10,000, and 0–100,000) prior to children receiving any corrective feedback from the experimenter. These children's results were then compared to the performance of college-aged adults who were asked to complete the same task. From the results of Experiment 1, we also hoped to identify an age group where there was a large potential for generalizing a linear representation of numbers to much larger orders of magnitude. Experiment 1 also allowed us to investigate whether the assumptions of Ebersbach et al.'s (2008) segmented linear model held at larger orders of magnitude (e.g., 0–10,000 and 0–100,000 scales).

In Experiment 2, we attempted to (a) determine how widely children generalized linear representations of numerical magnitudes after they were given feedback on the correctness of their estimates on a smaller scale and (b) determine whether alignment of small and large numeric scales was sufficient to produce generalization to larger numerical contexts. To meet our goals for Experiment 2, we brought number lines for large scales (0–1,000, 0–10,000, and 0–100,000) into progressive alignment with scales that children already represented linearly (0–100). The alignment procedure we used is considered "progressive" in two senses. First, all children's comparisons were supported by increasing perceptual similarity of small (training problems) and large scales (generalization problems) by matching color of units and orders of magnitude. Second, in our progressive alignment and multiple exemplars conditions, children were able to compare 0–100 problems with 0–1,000, 0–10,000, and 0–100,000 problems; in contrast, children in the no alignment condition were not given this opportunity. Finally, to test for representational change, we examined numerical estimates on a 0–1,000 posttest, where children were not given feedback, perceptual support, or the opportunity to compare problems to smaller scales.

Experiment 1: Long-Term Changes in Representations of Large Numerical Magnitudes

In Experiment 1, we investigated long-term changes in children's estimates of large numerical magnitudes (e.g., 940, 9,400, and 94,000) by examining estimates of second graders, third graders, sixth graders, and adults on 0–1,000, 0–10,000, and 0–100,000 number lines. Previously, number line estimation results have shown parallel developmental changes at different ages. Thus, a logarithmic-to-linear shift occurs between kindergarten and second grade on 0–100 number lines (Siegler & Booth, 2004), and between second and fourth grade for estimates on 0–1,000 number lines (Siegler & Opfer, 2003). In this study, we examined whether yet another logarithmic-to-linear shift occurs among children older than second graders on 0–10,000 and 0–100,000 number lines. This issue was important because findings of parallel developmental changes at older ages would suggest that for many years children face the problem of not realizing when to scale up their linear representation and would thereby raise the scientific issue of how they might make that realization.

Method

Participants. Participants were 24 second graders (mean age = 7.9, *SD* = 0.33; 14 girls and 10 boys), 24 third graders (mean age = 9.23, *SD* = 0.36; 11 girls and 13 boys), 24 sixth graders (mean age = 12.06, *SD* = 0.36; 12 girls and 12 boys), and 24 college-aged adults (mean age = 19.96, *SD* = 1.8; 16 women and 8 men). Children were recruited from elementary schools in largely European American, middle-class suburbs surrounding two large metropolitan cities in the United States. Adults were recruited from an introductory psychology course at a large university in the same city. Children participated in return for a small prize (e.g., a sticker), whereas adults received credit toward their introductory psychology course. One female graduate student and one female research assistant served as experimenters.

Design and procedure. Children and adults were given a number line estimation task that consisted of a line flanked by two hatch marks where the left hatch mark was labeled ''0,'' and right hatch mark was labeled with either ''1,000,'' ''10,000,'' or ''100,000'' depending on experimental condition (see Figure 1). Magnitude of overall scale (0–1,000, 0–10,000, and 0–100,000) was a between-subjects variable to guard against order effects. Across all conditions, participants were asked to estimate the position of a third number that appeared above the midpoint of the number line by making a hatch mark through the line. To-be-estimated numbers were chosen to maximize discriminability of logarithmic and linear functions (i.e., by oversampling at the low end of the scale) and to minimize influence of specific knowledge (e.g., 500 is halfway between 0 and 1,000).

Participants were presented with 1 number line estimation problem per page, which ensured that participants were unable to reference their previous estimates. Participants from each age group were randomly assigned to one of three conditions where they completed 10 number line estimation problems, presented in random order and without feedback from the experimenter: (a) in the 0–1,000 condition, participants estimated the magnitudes 20, 50, 80, 110, 150, 250, 490, 610, 730, and 940; (b) in the 0–10,000 condition, participants estimated the magnitudes 200, 500, 800, 1,100, 1,500, 2,500, 4,900, 6,100, 7,300, and 9,400; and (c) in the 0–100,000 condition participants estimated the magnitudes 2,000, 5,000, 8,000, 11,000, 15,000, 25,000, 49,000, 61,000, 73,000, and 94,000.

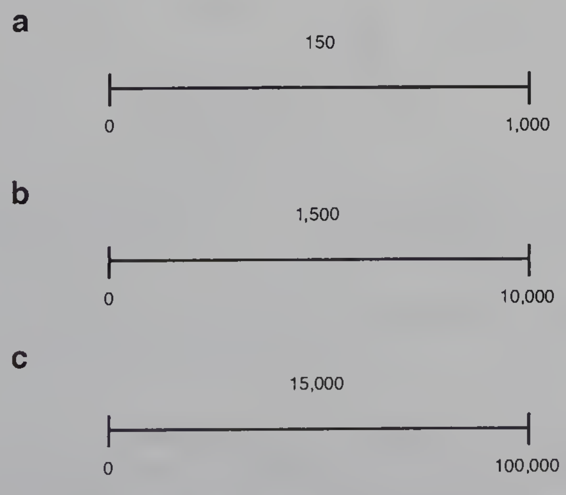


Figure 1. Experiment 1: Illustration of stimuli: (a) 0–1,000 scale, (b) 0–10,000 scale, and (c) 0–100,000 scale.

Results and Discussion

We first examined accuracy of numerical estimates across three experimental conditions. To measure accuracy, we first converted the participant's hatch mark to a numeric value by taking the proportion of the line indicated by the participant (i.e., linear distance from ''0'' mark to participant's hatch mark, divided by total length of line) and then multiplied this proportion by the right numerical anchor (1,000, 10,000, or 100,000). Then, percent absolute error of each participant's error (0%–100%) was calculated by taking the mean absolute difference between each of the participants' estimated values and the actual values divided by the total scale (1,000, 10,000, or 100,000). Finally, accuracy scores were computed by subtracting percent absolute error from 100%.

To analyze long-term changes in estimation accuracy, we conducted a 4 (age group: second graders, third graders, sixth graders, adults) × 3 (scale: 1,000, 10,000, 100,000) analysis of variance (ANOVA) on accuracy scores. As expected, accuracy increased with age, $F(3, 95) = 27.43$, $p < .0001$, $\eta^2 = .74$, and decreased with scale, $F(2, 95) = 7.51$, $p < .001$, $\eta^2 = .13$. These two main effects were also qualified by a significant Age Group × Scale interaction, $F(6, 95) = 2.38$, $p < .05$, $\eta^2 = .13$ (see Table 1). Although third and sixth graders outperformed second graders on 0–1,000 and 0–10,000 number lines, third and sixth graders failed to outperform second graders on 0–100,000 number lines. Thus, age group differences in estimation accuracy depended on numerical context (cf. Siegler & Opfer, 2003), with larger age differences present on 0–1,000 and 0–10,000 number lines than were present on 0–100,000 number lines.

To examine whether age differences in estimation accuracy for 0–10,000 and 0–100,000 number lines were linked to the logarithmic-to-linear shift seen in children's estimates on 0–1,000 number lines (Opfer & Siegler, 2007; Opfer & Thompson, 2008; Siegler & Opfer, 2003; Thompson & Opfer, 2008) and 0–100 number lines (Booth & Siegler, 2006, 2008; Laski & Siegler, 2007; Siegler & Booth, 2004; Siegler & Mu, 2008), we regressed participants' median estimates against the to-be-estimated number, and we compared the fit of the best fitting linear and logarithmic regression functions across the three age groups and scales (see Figure 2). Consistent with previous usage (e.g., Siegler & Opfer, 2003; Thompson & Opfer,

Table 1
Accuracy of Numerical Estimates by Age Group and Condition

	0–1,000	0–10,000	0–100,000
Age	$F(3, 31) = 10.92, p < .0001$ Adults, $M = 96.2\% >$ Sixth, $M = 91\%$, $d = 1.36$ Adults, $M = 96.2\% >$ Third $M = 90.5\%$, $d = 1.12$ Adults, $M = 96.2\% >$ Second, $M = 80.3\%$, $d = 2.94$ Sixth, $M = 91\% >$ Second, $M = 80.3\%$, $d = 1.72$ Third, $M = 90.5\% >$ Second, $M = 80.3\%$, $d = 1.45$	$F(3, 31) = 28.96, p < .0001$ Adults, $M = 96.7\% >$ Sixth, $M = 89.2\%$, $d = 1.42$ Adults, $M = 96.7\% >$ Third, $M = 85.6\%$, $d = 1.93$ Adults, $M = 96.7\% >$ Second, $M = 67.9\%$, $d = 5.91$ Sixth, $M = 89.2\% >$ Second, $M = 67.9\%$, $d = 3.03$ Third, $M = 85.6\% >$ Second, $M = 67.9\%$, $d = 2.39$	$F(3, 31) = 6.36, p < .01$ Adults, $M = 96.3\% >$ Sixth, $M = 80.4\%$, $d = 1.84$ Adults, $M = 96.3\% >$ Third, $M = 77.7\%$, $d = 2.65$ Adults, $M = 96.3\% >$ Second, $M = 73.8\%$, $d = 1.96$

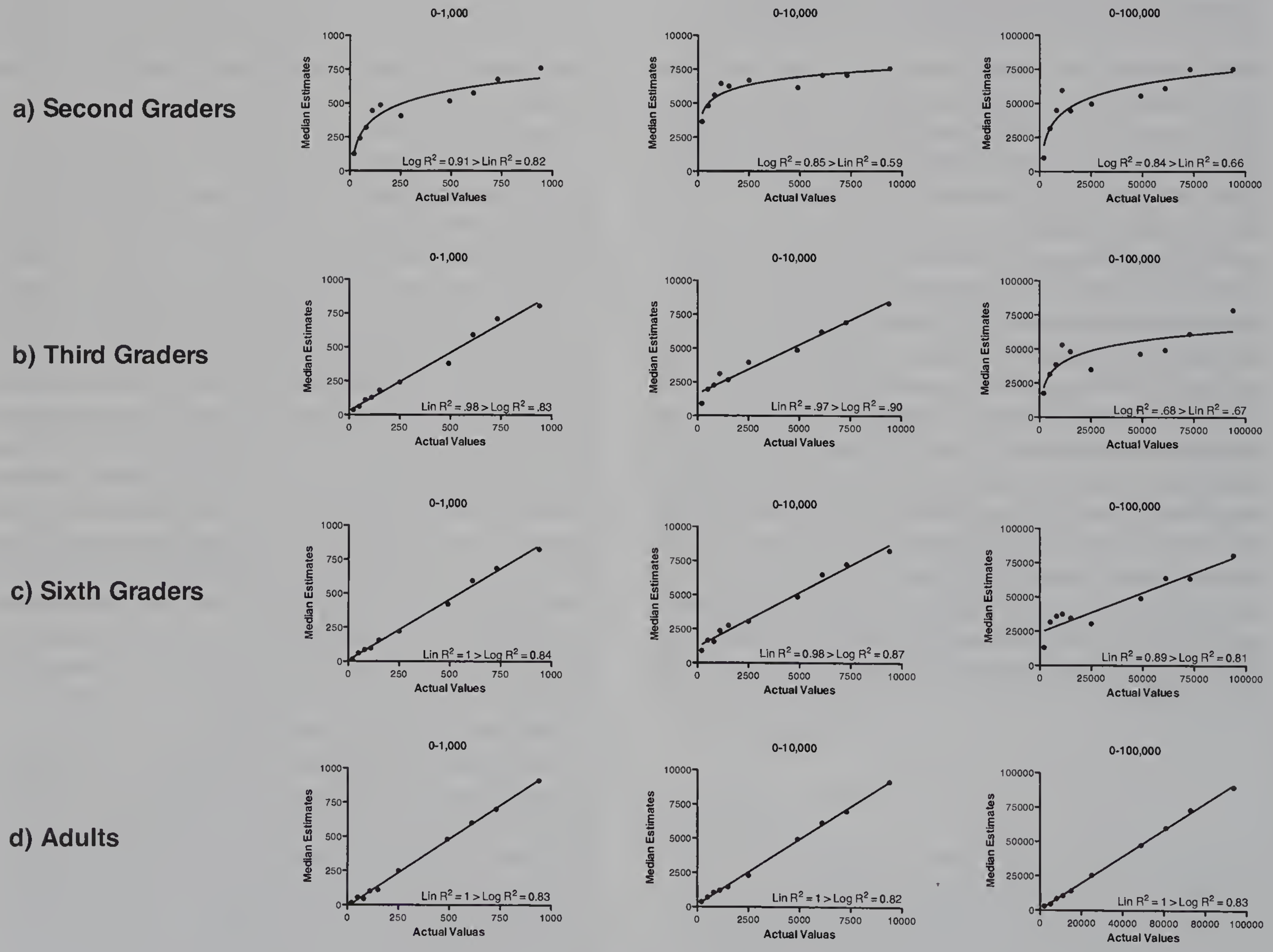


Figure 2. Experiment 1: Median estimates of second graders, third graders, sixth graders, and adults on 0–1,000, 0–10,000, and 0–100,000 number lines.

2008), median estimates were used to minimize the impact of potential outliers at the group level (although similar logarithmic and linear fits were obtained when we used mean estimates in Experiments 1 and 2). Across all three scales, second graders' median estimates were better fit by logarithmic than linear regression functions (0–1,000: logarithmic $R^2 = .91 >$ linear $R^2 = .82$; 0–10,000:

logarithmic $R^2 = .85 >$ linear $R^2 = .59$; 0–100,000: logarithmic $R^2 = .84 >$ linear $R^2 = .66$). Third graders' median estimates, however, were better fit by linear than logarithmic functions on 0–1,000 and 0–10,000 scales (0–1,000: linear $R^2 = .98 >$ logarithmic $R^2 = .83$; 0–10,000: linear $R^2 = .97 >$ logarithmic $R^2 = .90$), but their median estimates on 0–100,000 number lines were better fit by logarithmic than by linear regression functions (0–100,000: logarithmic $R^2 = .68 >$ linear $R^2 = .67$). Finally, median estimates on all three number lines were better fit by linear than by logarithmic regression functions for sixth graders (0–1,000: linear $R^2 = 1 >$ logarithmic $R^2 = .84$; 0–10,000: linear $R^2 = .98 >$ logarithmic $R^2 = .87$; 0–100,000: linear $R^2 = .89 >$ logarithmic $R^2 = .81$) and adults (0–1,000: linear $R^2 = 1 >$ logarithmic $R^2 = .83$; 0–10,000: linear $R^2 = 1 >$ logarithmic $R^2 = .82$; 0–100,000: linear $R^2 = 1 >$ logarithmic $R^2 = .83$).

To determine if the results at the group level held at the individual level, we next regressed each individual participant's estimates against the actual number using logarithmic and linear regression functions, and we used logistic regression to examine the odds of providing linear series of estimates as a function of age (see Figure 3). On 0–1,000 number lines, odds of generating linear estimates tended to increase with age, $\hat{\beta} = .72$, $z = 1.90$, Wald(1, $N = 32$) = 3.60, $p = .058$, indicating that with each year of age children were 2.05 times as likely to generate linear series of estimates. On 0–10,000 number lines, odds of generating linear estimates also increased with age, $\hat{\beta} = .68$, $z = 2.28$, Wald(1, $N = 32$) = 5.2, $p < .05$, indicating that with each year of age children were 1.97 times as likely to generate linear series of estimates. Finally, on 0–100,000 number lines, odds of generating linear estimates increased with age, $\hat{\beta} = .35$, $z = 2.32$, Wald(1, $N = 32$) = 5.4, $p < .05$, indicating that with each year of age children were 1.41 times as likely to generate linear series of estimates.

In summary, estimation accuracy increased with age, with age differences in accuracy being smaller on progressively larger numerical scales. The same pattern was also evident in the proportion of children producing linear series of estimates. Within each numerical scale, generation of linear series of estimates increased with age, but the overall proportion of children generating linear series of estimates again decreased with progressively larger scales. Thus, taken with previous results on numerical estimation in 0–10 and 0–100 contexts, the results of Experiment 1 point to a problem in numerical estimation that is faced to some extent by children from preschool to early adolescence: How to recognize that the linear representation used for small numbers is also appropriate to use for large numbers?

Experiment 2: Effect of Progressive Alignment on Representational Change

In Experiment 2, we investigated the effect of progressive alignment on children's application of linear representations to large numerical scales. Results of Experiment 1 suggested that second graders would serve as ideal participants: Second graders typically represent numbers as increasing linearly on 0–100 number lines and logarithmically on 0–1,000 number lines (Siegler & Booth, 2004; Siegler & Opfer, 2003), thereby making an analogy between the two scales potentially effective in improving representations for larger numbers. To test this idea, we

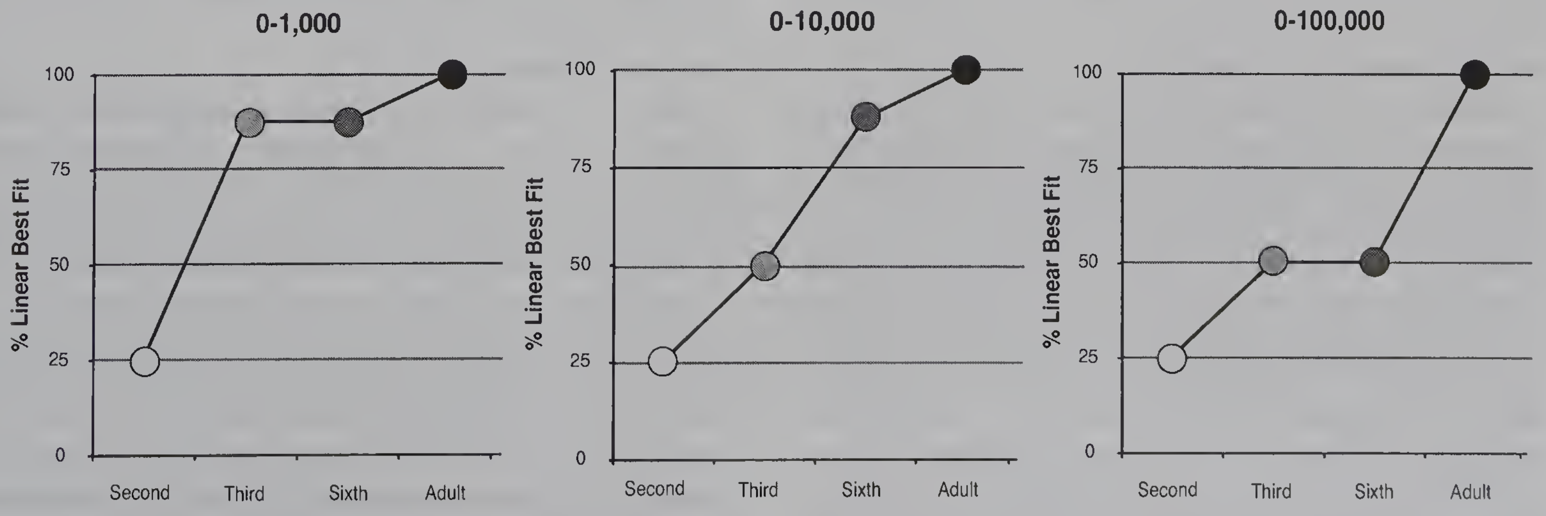


Figure 3. Experiment 1: Percentage of second graders, third graders, sixth graders, and adults who were best fit by the linear regression function on 0–1,000, 0–10,000, and 0–100,000 number lines.

examined effects of three training procedures to elicit linear estimates from second graders on 0–1,000, 0–10,000, and 0–100,000 number lines. In all three procedures, children were given feedback when making estimates on 0–100 number lines. The critical difference between procedures, however, was whether and how 0–100 number lines were aligned with 0–1,000, 0–10,000, and 0–100,000 number lines. Our central hypothesis was that progressive alignment of small and large number lines would lead children to change their estimates of numerical magnitude on large number lines, despite being given no direct feedback that they should do so.

We hypothesized that progressive alignment would promote representational change in the following way. Participants should first make surface-level comparisons across 0–100 training problems and note that a particular numerosity should be placed in the same location on the number line regardless of whether the numerosity is presented in the context of a blank 0–100 number line, a 0–100 pear line, a 0–100 cherry line, or a 0–100 carrot line. Then, we hypothesized that participants should note color of units (e.g., pears, cherries, carrots) corresponding to differing orders of magnitude (0–1,000 with one green zero, 0–10,000 with two red zeros, and 0–100,000 with three orange zeros, respectively), and this should subsequently highlight the underlying decimal system (15 cherries:100 cherries::1,500:10,000).

Method

Participants. Participants were 46 second graders (mean age = 7.78, *SD* = 0.41) recruited from elementary schools in largely European American, middle-class suburbs surrounding two large metropolitan cities in the United States. There were 20 girls and 26 boys that participated in return for a small prize (e.g., a sticker). One female graduate student served as experimenter.

Design and procedure. Children estimated the placement of numbers on number lines across three phases of the experiment: training, generalization, and posttest (see Figure 4). Number lines were flanked by two hatch marks, the left hatch mark was labeled "0," and right hatch mark was labeled "100," "1,000," "10,000," or "100,000." On each trial, children estimated the position of a number (one per number line) by making a hatch mark through the line. To-be-estimated numbers (2, 5, 8, 11, 15, 25, 49, 61, 73, 94, and multiples thereof) were chosen to reduce influence of specific knowledge (e.g., that 50 is half of 100) and to oversample the low end of the scale to maximize discriminability of logarithmic and linear functions.

Figure 4 illustrates training, generalization, and posttest phases for three experimental conditions: no alignment, multiple exemplars, and progressive alignment. During training (Figure 4, left column), all participants received corrective feedback in the 0–100 scale that indicated how close to or far from the actual location of the to-be-estimated number their hatch marks were to ensure children possessed a linear representation in this scale (see Opfer & Siegler, 2007, for a more detailed description of the feedback procedure). We reasoned that if children possessed a linear representation in the 0–100 scale, participants across all experimental groups should be equally likely to bootstrap their linear representation to larger numerical scales (0–1,000, 0–10,000, and 0–100,000).

All training problems specified units (i.e., green pears, red cherries, and orange carrots) and were completed in the following order: blank number line, pear number line, cherry number line, then carrot number line. After children completed these training problems, they were told that their estimates did not differ much over different units (e.g., "It does not matter if you see pears, cherries, carrots, or nothing at all after the number 15, if you see the number 15, you should make your mark right here").

In the generalization phase (Figure 4, middle column), we highlighted similarity of generalization and training problems (that color of units matched color of zeros in right numerical anchor) and told participants to "try some more problems just like the ones you just finished" as the experimenter drew an arrow from the relevant training problem to its related generalization problem. During generalization, participants made estimates in the 0–100, 0–1,000, 0–10,000, and 0–100,000 scales without corrective feedback for the numerosity on which they were just trained.

Figure 4 illustrates training and generalization for the number 15. Participants completed one training problem at a time (e.g., 15 on a 0–100 number line, 15 pears on a 0–100 pear line, 15 cherries on a 0–100 cherry line, and 15 carrots on a 0–100 carrot line), and then completed generalization problems in the 0–100 through 0–100,000 scale (e.g., 15 on a 0–100 number line, 150 on a 0–1,000 number line, 1,500 on a 0–10,000 number line, and 15,000 on a 0–100,000 number line). This training and generalization procedure was repeated with the other nine numerosities we investigated for a total of 40 training trials and 40 generalization trials.

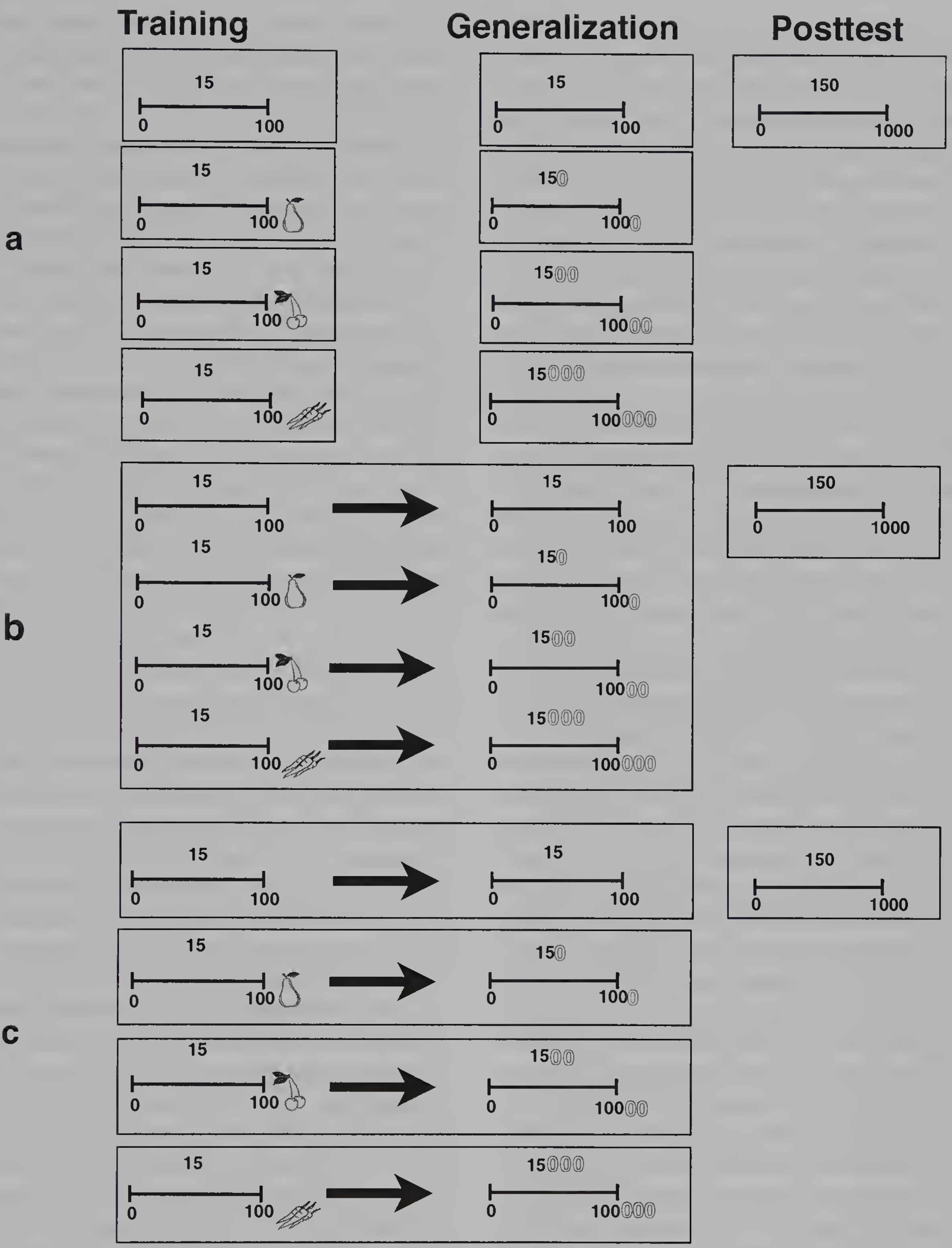


Figure 4. Experiment 2: Illustration of stimuli: (a) no alignment condition, (b) multiple exemplars condition, and (c) progressive alignment condition.

Note. Units in 0–100 training problems (i.e., pears, cherries, and carrots) were colored to correspond to the final 1–3 zeros in 0–1,000, 0–10,000, and 0–100,000 generalization problems (e.g., 3 carrots and the final three zeros in 100,000 were colored orange). Participants in the no alignment condition were presented one training or generalization problem per page, so these participants could not compare across training and generalization problems. Participants in the multiple exemplars condition were presented all training and generalization problems on one page, so these participants could compare training to other training problems, generalization to other generalization problems, or training to generalization problems. Participants in the progressive alignment condition were presented with one training–generalization pair at a time, so these participants could compare across relevant training–generalization pairs only.

Following completion of the training and generalization procedure detailed earlier, participants completed a 10-problem posttest phase (Figure 4, right column). Participants completed the same 10 problems administered for the 0–1,000 generalization phase (e.g., 20, 50, 80, 110, 150, 250, 490, 610, 730, 940). Like the generalization phase, participants did not receive corrective feedback during posttest. Unlike generalization, posttest problems contained no perceptual support (e.g., fruit and vegetable icons and colored zeros).

To test the effect of alignment on transfer from training to generalization and posttest problems, participants were randomly assigned to one of three between-subjects experimental conditions: no alignment, multiple exemplars, and progressive alignment (cf. Figures 4a to 4c). In the no alignment condition, participants received training and generalization problems one at a time; thus, participants were not able to compare problems on which they received feedback during training (e.g., placement of 15 cherries on a 0–100 cherry line) with problems they completed during the generalization phase (e.g., 1,500 on a 0–10,000 number line). In the multiple exemplars and progressive alignment conditions, participants received the same training problems and corrective feedback as in the no alignment condition, but generalization problems were presented alongside previously solved training problems, thereby allowing children in these groups to compare generalization problems to training problems. The multiple exemplars and progressive alignment conditions differed from one another in that participants in the progressive alignment group were *only* able to compare one training–generalization pair at a time, thereby focusing on the critical similarity (e.g., 15 cherries:100 cherries::1,500:10,000). In contrast, participants in the multiple exemplars condition saw *all* training and generalization problems on one page. Participants in the multiple exemplars condition were able to compare training problems to generalization problems just like participants in the progressive alignment condition (and thus draw the relevant analogy), but participants in the multiple exemplars condition were also given the opportunity to make less useful comparisons, such as comparing training problems to training problems (cherries vs. carrots) and comparing generalization problems to generalization problems (10,000 vs. 100,000). To put it another way, participants in the progressive alignment condition were constrained in the types of comparisons they could make, whereas participants in the multiple exemplars condition were unconstrained in the comparisons they could make.

Results and Discussion

In the following sections, we will analyze children's number line estimates during training, generalization, and at posttest to determine whether children were capable of using an analogy to scale up their linear representation from 0–100 number lines to larger numerical magnitude contexts.

Training. We first examined accuracy of numerical estimates during training to confirm that all three experimental groups performed well on 0–100 number lines, which were problems from which the groups were expected to generalize in the generalization and posttest phases. As expected, a 3 (condition: no alignment, multiple exemplars, progressive alignment) × 4 (unit type: none, pear, cherry, carrot) ANOVA revealed no effect of condition or Condition × Unit Type interaction. Presumably due to the fixed order of items, however, there was a small but reliable effect of unit type, $F(3, 129) = 15.13$, $p < .001$, $\eta^2 = .24$, with less accurate estimates for the first type of units presented (none, $M = 89\%$) than for the other three types of units (pear: $M = 94\%$, $d = 0.82$; cherry: $M = 94\%$, $d = 1.21$; carrot: $M = 93\%$, $d = 1.09$; $ps < .01$), which did not differ significantly.

To confirm that these high levels of accuracy were associated with use of linear representations, we again compared the fit of the best fitting linear and logarithmic regression functions to median numerical estimates across three experimental conditions. Across all training problems, participants in the no alignment condition produced more linear than logarithmic series of estimates (blank: linear $R^2 = .99 >$ logarithmic $R^2 = .85$; pear: linear $R^2 = .99 >$ logarithmic $R^2 = .86$; cherry: linear $R^2 = 1 >$ logarithmic $R^2 = .85$; carrot: linear $R^2 = 1 >$ logarithmic $R^2 = .85$), as did participants in the multiple exemplars condition (blank: linear $R^2 = .99 >$ logarithmic $R^2 = .86$; pear: linear $R^2 = 1 >$ logarithmic $R^2 = .84$; cherry: linear $R^2 = 1 >$ logarithmic $R^2 = .84$; carrot: linear $R^2 = 1 >$ logarithmic $R^2 = .84$), and the progressive alignment condition (blank: linear $R^2 = .873 >$ logarithmic $R^2 = .869$; pear: linear $R^2 = 1 >$ logarithmic $R^2 = .84$; cherry: linear $R^2 = 1 >$ logarithmic $R^2 = .83$; carrot: linear $R^2 = 1 >$ logarithmic $R^2 = .83$). A similar pattern also emerged from the examination of individual children's estimates. Regardless of units, most children's estimates were best fit by the linear function in the no alignment condition ($M = 93\%$ best fit by linear), the multiple exemplars condition ($M = 88\%$ best fit by linear), and the progressive alignment condition ($M = 86\%$ best fit by linear). Further, tests of Fisher's exact probabilities revealed no association between experimental group and best fitting regression function (logarithmic versus linear; $ps > .05$).

Thus, all three experimental groups were likely to use an accurate, linear representation of numerical magnitudes in the 0–100 scale, suggesting they were equally likely to benefit from applying their linear representations of the 0–100 scale to 0–1,000, 0–10,000, and 0–100,000 scales. In the next section, we explored whether all three groups did in fact apply their linear representations to larger scales and whether that application was fostered by alignment of small and large numerical scales.

Generalization. To test the effect of progressive alignment on generalization of linear representations to larger scales, we first examined accuracy across three experimental groups. To do this, we conducted a 3 (condition: no alignment, multiple exemplars, progressive alignment) × 4 (scale: 0–100, 0–1,000, 0–10,000, 0–100,000) repeated measures ANOVA on accuracy scores on generalization problems. As expected, there was a main effect of scale, $F(3, 41) = 28.73$, $p < .001$, $\eta^2 = .02$, and a main effect of condition, $F(2, 43) = 10.44$, $p < .001$, $\eta^2 = .91$, although these main effects were qualified by a significant Condition × Scale interaction, $F(6, 84) = 2.99$, $p < .01$, $\eta^2 = .04$.

Further, to investigate significant condition differences in levels of accuracy across generalization problems, we conducted post hoc ANOVAs to determine how differing amounts of alignment (e.g., no alignment, multiple exemplars, progressive alignment) affected accuracy of children's estimates (see Table 2). As anticipated, there were no differences in accuracy across experimental conditions for the 0–100 blank number line problems, presumably because second graders typically represented magnitudes of these numbers as increasing linearly. However, children assigned to the multiple exemplars and progressive alignment conditions produce more accurate series of number line estimates for 0–1,000, 0–10,000, and 0–100,000 generalization problems as compared to those children assigned to the no alignment condition.

To assess whether these changes in accuracy were caused by the logarithmic-to-linear shift, we next compared the fit of the best-fitting linear and logarithmic functions to median numerical estimates across three experimental groups (Figure 5). As in performance on 0–100 problems during training, the best fitting function for all experimental conditions on the 0–100 generalization problems was the linear rather than logarithmic regression function (no alignment, linear $R^2 = 1.0 >$ logarithmic $R^2 = .86$; multiple exemplars, linear $R^2 = .99 >$ logarithmic $R^2 = .85$; progressive alignment, linear $R^2 = .99 >$ logarithmic $R^2 = .85$).

Experimental groups differed, however, for larger generalization problems (0–1,000, 0–10,000, and 0–100,000). For the no alignment group, the best fitting function was logarithmic across the 0–1,000 and 0–10,000 generalization problems (logarithmic $R^2 = .9$ and .85, respectively); on 0–100,000 generalization problems, both the linear ($R^2 = .46$) and logarithmic function ($R^2 = .39$) provided uncommonly poor fits to children's estimates. For the multiple exemplars group, the best fitting function was the linear one for both 0–1,000 and 0–10,000 generalization problems (linear $R^2 = .96$ and .84, respectively), but the logarithmic function was the best fitting function across 0–100,000 generalization problems (logarithmic $R^2 = .86$). Finally, for the progressive alignment group, the linear function provided the best fit across 0–1,000, 0–10,000, and 0–100,000 generalization problems (linear $R^2 = 1.00$, .98, and .98, respectively).

To ensure that these condition differences in model fit were due to the effect of alignment and not due to irregularities arising from averaging, we used Fisher exact probability tests to examine reliability of association between condition and percentage of individual children best fit by the linear function. On the two smaller scales, 0–100 and 0–1,000, we observed no significant associations between condition and fit by the linear function (on

Table 2
Accuracy of Numerical Estimates by Scale and Condition

	0–1,000	0–10,000	0–100,000
Condition	$F(2, 45) = 8.38$, $p < .001$ Progressive, $M = 90\% >$ No, $M = 78\%$, $d = 1.25$ Multiple exemplars, $M = 89\% >$ No, $M = 78\%$, $d = 1.09$	$F(2, 45) = 6.06$, $p < .01$ Progressive, $M = 83\% >$ No, $M = 63\%$, $d = 0.95$ Multiple exemplars, $M = 84\% >$ No, $M = 63\%$, $d = 1.16$	$F(2, 45) = 10.89$, $p < .001$ Progressive, $M = 85\% >$ No, $M = 63\%$, $d = 1.48$ Multiple exemplars, $M = 82\% >$ No, $M = 63\%$, $d = 1.43$

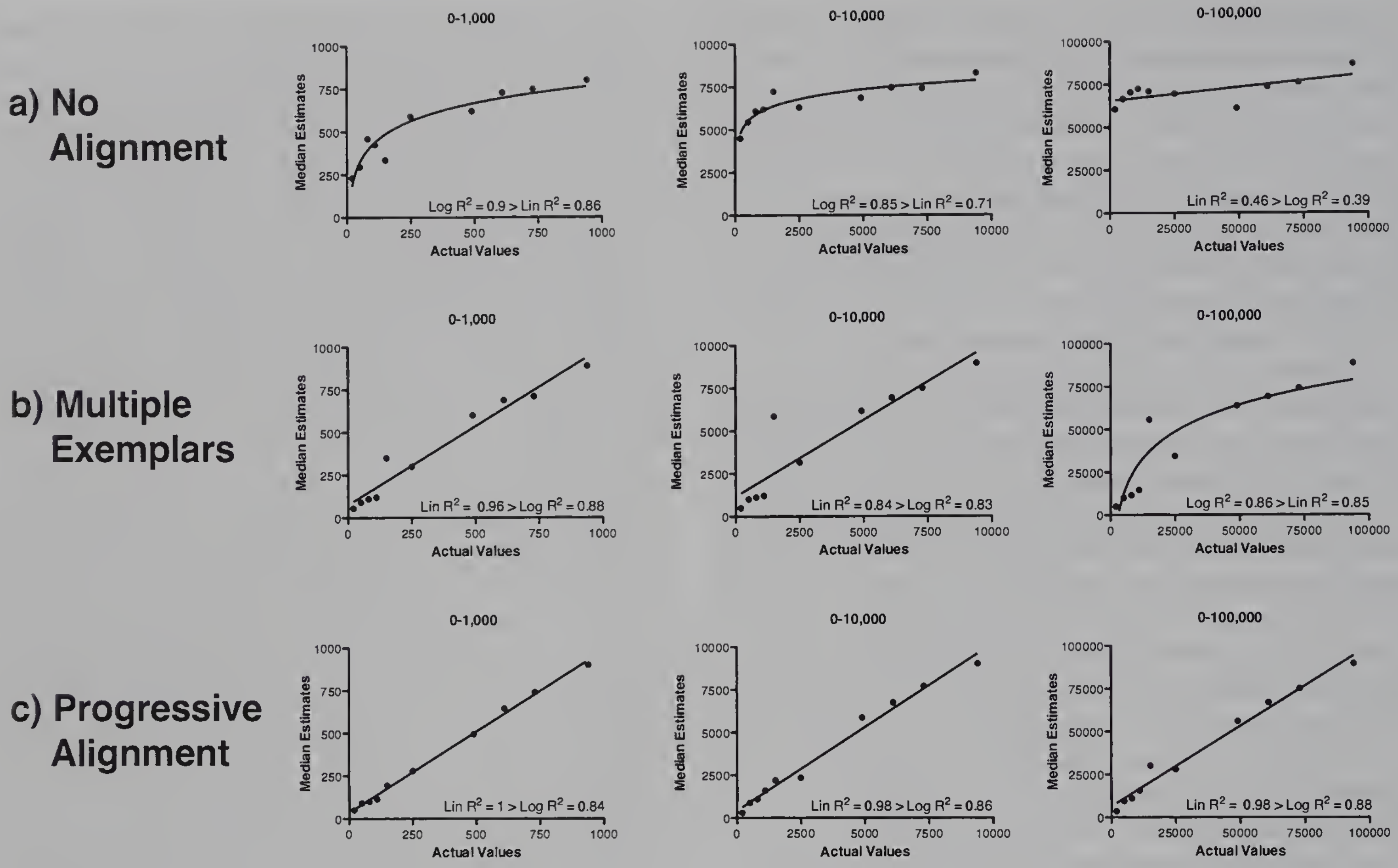


Figure 5. Experiment 2 (generalization): Median estimates of participants in the no alignment, multiple exemplars, and progressive alignment groups on 0–1,000, 0–10,000, and 0–100,000 number lines.

0–100 scales: no alignment, 93%; multiple exemplars, 100%; progressive alignment, 100%; on 0–1,000 scales: no alignment, 57%; multiple exemplars, 69%; progressive alignment, 81%). On the two larger scales, 0–10,000 and 0–100,000, however, we observed significant associations between condition and fit by the linear function, with the progressive alignment group outperforming the other two experimental groups (on 0–10,000 scales: no alignment, 50%; multiple exemplars, 56%; progressive alignment, 100%; on 0–100,000 scales: no alignment, 50%; multiple exemplars, 50%; progressive alignment, 100%; *ps* < .05; compare to results of Experiment, 1 illustrated in Figure 3). Thus, while children generalized linear representations of numerical magnitude to 0–1,000 scales with and without the help of alignment of small and large numerical scales, children's generalization of linear representations to 0–10,000 and 0–100,000 scales only occurred when these scales were individually aligned with 0–100 scales, as they were for children in the progressive alignment group.

In summary, findings from the generalization phase of Experiment 2—greater estimation accuracy on number lines with larger numeric values, better fits of the linear regression function to estimates on number lines with larger numeric values, and strong associations between alignment and likelihood of providing linear estimates on number lines with larger numeric values—indicate that progressive alignment of large and small number lines led children to scale up their linear representations of numerical magnitude more so than no alignment of large and small scales (no alignment condition) or alignment of units as well as alignment of large and small scales (multiple exemplars condition).

We speculate the reason children did not scale up their linear representation in the multiple exemplars condition, which presumably offered a wealth of alignment possibilities, may have resulted from learners failing to align the most effective pairs (e.g., 0–100 pears training problem compared to 0–1,000 generalization problem with one green zero). That is, in the multiple exemplars condition, children were not constrained in the types of comparisons they made across problem types because they saw all problems on one page. When faced with the ability to compare any of the number lines on the page, learners may have failed to notice surface similarities across their initial comparisons.

If the children failed to see how the problems encompassing small and large numeric scales related to one another, they may have given up on their attempts to find commonalities. Alternatively, children may have used a more shallow comparison process by simply scanning the multiple exemplars for obvious, surface commonalities as compared to engaging in a deeper comparison process that might highlight structural similarities (e.g., decimal system) across the small and large numerical scales. In the next section, we examined whether advantages of progressive alignment over other training procedures persisted to a posttest, where alignment of large and small scales desisted.

Posttest. To determine whether participants had transferred their linear representation of number to larger numbers generally or just to number lines with perceptual support (e.g., fruit and vegetable icons and colored zeros), we examined performance on a blank 0–1,000 number line that was unaccompanied by previously solved problems and without units and zeroes sharing a common color. Our reasoning was that if children had grasped the analogy we intended, they would distance their hatch marks as a linear function of actual value on number lines identical to those used in previous studies (i.e., without perceptual support).

As illustrated in Figure 6, estimates of participants in the progressive alignment group were indeed better fit by the linear function ($R^2 = .95$) than by the logarithmic ($R^2 = .76$). In contrast, children's estimates in the no alignment group were better fit by the logarithmic function ($R^2 = .91$) than by the linear ($R^2 = .66$), and children's estimates in the multiple exemplars group were also better fit by the logarithmic function ($R^2 = .83$) than by the linear ($R^2 = .58$). Finally, the same advantage of the progressive alignment group was evident in the percentage of children best fit by the linear as opposed to the logarithmic function (compare with 0–1,000 condition in Figure 3), with 63% of children in the progressive alignment group generating linear series of estimates versus 29% of children in the no alignment group, $\chi^2(1) = 3.45$, $p = .06$, and 25% of children in the multiple exemplars group, $\chi^2(1) = 4.57$, $p < .05$.

General Discussion

The ability to carry out effortless structural alignment is a hallmark of human cognitive processing (Gentner & Markman, 1997; although see Kurtz et al., 2001, for an instance of more effortful alignment) and a potentially general mechanism of cognitive change (Chen & Klahr, 1999; Gentner et al., 2001; Holyoak & Thagard, 1995; Opfer & Siegler, 2004). In our study, we investigated whether structural alignment of small numerical scales and large numerical scales might lead to a representational change in the domain of number (logarithmic-to-linear shift) that has been observed across a large number of age groups (Siegler & Booth, 2004; Opfer & Thompson, 2008; Siegler & Opfer, 2003).

Our first experiment illustrated large numeric contexts in which second-, third-, and sixth-grade children produced linear versus logarithmic series of numerical magnitude estimates. Our results supported the idea that abandoning logarithmic representations of numerical magnitude is not an all-or-none process. Rather, extending linear representations of number to ever-larger numeric scales occurs gradually over time, with half or more of each group of children continuing to use logarithmic representations on one or more tasks (e.g., 50% of sixth graders on 0–100,000; 50% of third graders on 0–10,000 and 0–100,000; and 75% of second graders on 0–1,000, 0–10,000, and 0–100,000). Thus, for large numeric scales, children's use of logarithmic representations appeared unchanged by their experiences—either due to the low frequency with which they are likely to receive any information

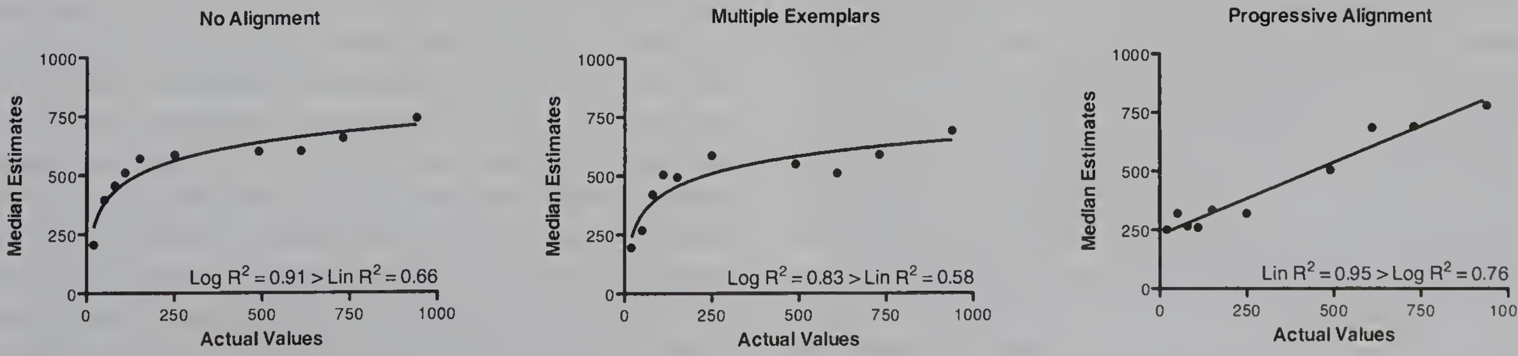


Figure 6. Experiment 2 (posttest): Median estimates of participants in the no alignment, multiple exemplars, and progressive alignment groups on unsupported 0–1,000 number lines.

about values in these large scales or due to an age-related conceptual limitation in representing large numeric values.

Our second experiment indicated that failure to abandon logarithmic representations of large numbers in favor of linear ones is likely due to poverty of input regarding large numerical values. Thus, when second graders—who produced logarithmic estimates at the highest rates (75%) across all three tasks in Experiment 1—were simply provided with the opportunity to align number line problems for small numerical scales (0–100) with number line problems of larger numerical scales (0–1,000, 0–10,000, and 0–100,000), children abandoned use of the logarithmic representation with surprising frequency. This representational change was most robust in the progressive alignment condition, where children could directly compare only relevant training and generalization problems and where 0% of children provided estimates better fit by the logarithmic than linear functions on 0–10,000 and 0–100,000 problems. Moreover, these children continued to generate linear estimates on posttest, which children in the multiple exemplars and no alignment conditions did with much less consistency. Progressive alignment of small and large numerical scales focused children's attention on the most informative surface level comparisons allowing children to search for deeper, structural commonalities as well. Children in the multiple exemplars condition were not at all constrained when attempting to compare small and large numerical scales. It is likely that surface dissimilarities (e.g., pears, cherries, carrots, 0–100, 0–1,000, 0–10,000, 0–100,000) confused learners in the multiple exemplars condition and prohibited these learners from discovering deeper, structural similarities across number lines of small and large numerical scales.

Analogy as a Mechanism of Representational Change

Why did alignment of small and large numeric scales produce rapid changes in numerical estimation and generalization of the linear representation? In our experiment, we aligned contexts in which children were familiar (e.g., 0–100) with larger, less familiar numeric contexts (e.g., 0–1,000 to 0–100,000). This alignment methodology apparently prompted second graders to scale up their linear representation of numbers by highlighting the underlying structure of the decimal system and thereby bootstrap the linear representation they already possessed in a familiar numeric context (0–100) to less familiar numeric contexts (0–1,000 through 0–100,000). Although this mechanism of representational change had been suggested in an earlier microgenetic study of numerical estimation (Opfer & Siegler, 2007), the present work provided unusually strong support for the idea by showing both that the representational change can occur without any feedback as well as by showing that the learning can be generalized very far outside the training space.

These findings on the effect of alignment in producing analogies are also apparent in a number of other empirical studies (Dixon & Dohn, 2003; Gentner, Loewenstein, & Thompson, 2003; Kotovsky & Gentner, 1996; Uttal, Schreiber, & DeLoache, 1995), where alignment has led people to compare examples and thereby extract an abstract relational schema that might be applied to other relevant problems. For example, novice negotiators who were learning about negotiation strategies were asked to compare written case studies in the hopes that a common negotiation strategy would be extracted despite differing surface features of presented problems. Novice negotiators who compared two written case studies detailing negotiation strategies were better able to abstract general problem-solving schemas which they later applied to future negotiation situations, like face-to-face negotiations, than were negotiators who were not explicitly told to compare cases or who received no case-based training whatsoever (Gentner et al., 2003; Loewenstein, Thompson, & Gentner, 2003). The general problem-solving schema was abstracted, or disembedded, from the original learning context so that it could be applied to other relevant case studies outside the original learning context.

Likewise, other investigators have similarly noted that alignment specifically and similarity comparison in general allows participants to disembed underlying structural relations in problem-solving tasks. In classic scale model problem-solving tasks conducted by DeLoache and her colleagues (e.g., DeLoache, Kolstad, & Anderson, 1991; Uttal et al., 1995), young children were shown where a hidden object could be found in a scale model and then were asked to locate the hidden object in the larger artificial room the model represented. Not only must children realize the scale model is related to the larger room and therefore locations in the model can be mapped directly onto this larger artificial room in a one-to-one fashion, but children must also hold in mind the representation of the scale model to find the hiding place of the object in the larger room (Uttal et al., 1995). According to DeLoache et al. (1991), one way that children successfully uncover the one-to-one mapping between the model and artificial room is by attending to overlapping similarity in the two spaces (e.g., object matches and size or scale of rooms). Further, Loewenstein and Gentner (2001) found that 3-year-olds' search behavior in the DeLoache hiding task was improved by their ability to compare two sequential hiding tasks transpiring across two highly similar hiding rooms whereas search behavior was not improved for those 3-year-olds who saw the highly similar hiding rooms but were not prompted to compare them.

Another example of participants' ability to abstract a general problem-solving schema through alignment is Dixon and Dohn's (2003) alternating sequence problems. In these problems, participants were shown a connected sequence of balance scales or gears and then were asked to predict the position of the last balance scale in sequence (up or down) or direction of movement of the last gear in sequence (clockwise or counterclockwise). Participants who discovered the underlying structural relation through comparison of multiple examples of alternating sequence problems often extracted the solution more quickly or readily and more consistently transferred the solution to new problem types than did those participants who were explicitly instructed on the relation uniting the problems (e.g., alternating sequences: up, down, up, down or clockwise, counterclockwise, clockwise, counterclockwise).

As in these cases of learning, children in our own experiments were given opportunities that seemed crucial for eliciting their analogies. First, they were given opportunity to compare across multiple training problems, much like novice negotiators comparing written case studies (Gentner et al., 2003), toddlers comparing scale models to larger rooms (DeLoache et al., 1991; Loewenstein & Gentner, 2001), adults comparing alternating sequence problems (Dixon & Dohn, 2003), and children comparing geometric shapes that differed in symmetry and monotonicity in match to sample tasks (Kotovsky & Gentner, 1996). Second, our participants were given perceptual matches to facilitate alignment of small number line problems (e.g., two red cherries in 0–100 problems) with large number line problems (e.g., two red zeros in 0–10,000 problems), just as DeLoache and colleagues' toddlers used high physical similarity across the size of the scale model and artificial room. Third, not only did presence of overlapping similarity lead participants to abstract structural relations, but children's superior understanding of the 0–100 scale was probably crucial, given the finding that increasing participants' knowledge base increases the likelihood of learners noticing structural relations (Kotovsky & Gentner, 1996). Thus, Experiment 2 appeared to provide many critical opportunities that children needed to draw analogies from more familiar to less familiar problems.

Alternative Role for Familiarity in Numerical Estimation: Segmented Linear Model

According to the segmented linear model, children's familiarity with numbers, as indexed by counting ability within a particular numeric scale, is a good predictor of whether children produce linear estimates (as in familiar numeric scales) or logarithmic estimates (as in unfamiliar numeric scales). Although, according to Ebersbach et al. (2008), the fit of the logarithmic function may be illusory, and children's estimates might be better fit by two linear regression functions that intersect at a "change point."

Results of Experiment 1 provided an unusual opportunity to evaluate this alternative model, and our results do not support the segmented linear model. Specifically, when we fit estimates using the segmented linear function, we found increasingly larger change points as numeric context increased from 0–1,000 (second grade: 158.7, third grade: 366.5, sixth grade: 386.3, adults: 421.35), 0–10,000 (second grade: 2,330, third grade: 2,500, sixth grade: 4,263, adults: 3,668), and 0–100,000 (second grade: 37,000, third grade: 26,592.5, sixth grade: 23,798, adults: 48,999.5), $r = .94$, $F(1, 11) = 70.89$, $p < .001$, suggesting that change point is a task-specific scaling parameter for a logarithmic function rather than an index of children's numeric familiarity. To test the idea that change point is just a task-specific scaling parameter, we next regressed mean age of each participant group (second grade: 7.9 years, third grade: 9.23 years, sixth grade: 12.06 years, adults: 19.96 years) against the segmented linear model's change point, and we found no correlation between age and supposed range of children's familiarity, $r = .15$, $F(1, 11) = .23$, $p > .05$, *ns*. Thus, although numeric anchor (0–1,000, 0–10,000, or 0–100,000) accounted for 94% of variation in change points, age, in comparison, accounted for an insignificant (15%) amount of variation in change points.

Although change point of the segmented linear model has been interpreted by Ebersbach et al. (2008) as reflecting a division in children's familiarity with numbers, the finding that change point varied directly with scale of the number line task (0–1,000 through 0–100,000) suggests that change point is better understood merely as a task-specific scaling parameter. This point is clearest for third graders, where data from both Ebersbach et al. (2008) and our own data indicate a change point of approximately 342 and 366 (respectively) on the 0–1,000 task. Taken alone, it is tempting to view 342–366 as an outer range where children's familiarity with numbers might end. However, since we found the same age group of children producing estimates with a change point of 2,500 and 26,592 on a 0–10,000 and 0–100,000 task (respectively) these findings clearly undercut this logic: Change point in a segmented linear model cannot simply index children's numeric familiarity. The alternative explanation—that participants' familiarity with numbers corresponds directly to the task to which they were randomly assigned—is clearly implausible. Thus, although participants in our studies appeared to use their knowledge of a familiar numeric scale (0–100) to draw an analogy to less-familiar scales (0–1,000 through 0–100,000), familiarity itself does not appear to provide a viable alternative explanation for the logarithmic pattern of estimates observed in Experiment 1.

Educational Implications

Understanding of place value, or interrelations of the numeric decimal system, is a part of children's blossoming number sense (U.S. Department of Education, 2008). This number sense, according to a recent report by the National Mathematics Advisory Panel "includes the ability to estimate results of computations and thereby to estimate orders of magnitude" (p. 18). The panel detailed the state of mathematics research and education in the United States and noted the importance of fostering children's number sense because "poor number sense interferes with learning algorithms and number facts and prevents use of strategies to verify if solutions to problems are reasonable" (p. 27).

Our current research on alignment of small and large numerical scales offers a possible means of providing students explicit as well as implicit instruction on place value and the decimal system. In Experiment 2, we showed that progressive alignment of small and large numerical scales leads young children to adopt a linear representation of number and abandon a logarithmic one, which lends support to important educational implications of our results. The educational implication of this alignment methodology is that it might be used as a simple classroom intervention that may allow children to being thinking about numbers in a more adult-like, linear fashion months and maybe even years before classical classroom instruction or life experiences would lead them to do so. As part of the simple intervention, teachers might align small numeric scales with larger ones and even provide a verbal or visual hint about how the decimal system works (e.g., crossing out the extra zero in 0–1,000 order of magnitude to show how it is similar to the 0–100 order of magnitude). To our knowledge, highlighting properties of the decimal system through alignment is not a technique commonly used by elementary school teachers, but based on previous evidence of causal and correlational links between numerical estimation and other mathematical skills (Booth & Siegler, 2006, 2008), we believe this intervention is likely to be an easy and effective one to implement.

References

Abdellatif, H. R., Cummings, R., & Maddux, C. D. (2008). Factors affecting the development of analogical reasoning in young children: A review of the literature. *Education, 129*, 239–249.

Alexander, P. A., Willson, V. L., White, C. S., & Fuqua, J. D. (1987). Analogical reasoning in young children. *Journal of Educational Psychology, 79*, 401–408.

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology Monograph, 102*, 353–376.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41*, 189–201.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016–1031.

Brannon, E. M. (2005). What animals know about numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). New York: Taylor & Francis.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition of the control of variables strategy. *Child Development, 70*, 1098–1120.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention and performance XXIV. Sensorimotor foundations of higher cognition* (pp. 527–574). Cambridge, MA: Harvard University Press.

Dehaene, S., Dehaene-Lambert, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*, 355–361.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43,* 1–29.

DeLoache, J. S., Kolstad, V., & Anderson, K. N. (1991). Physical similarity and young children's understanding of scale models. *Child Development, 62,* 111–126.

Dixon, J. A., & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition, 31,* 1082–1093.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115,* 1–43.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year-old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99,* 1–17.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8,* 307–314.

Furlong, E. E., & Opfer, J. E. (2009). Cognitive constraints on how economic rewards affect cooperation. *Psychological Science, 20,* 11–16.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., et al. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences, 6,* 3–40.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8,* 285–307.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393–408.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56.

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14,* 487–513.

Goswami, U., & Brown, A. (1990). Melting chocolate and melting snowman. Analogical reasoning and causal relations. *Cognition, 35,* 69–95.

Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press/Bradford Books.

Holyoak, K., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child Development, 55,* 2042–2055.

Holyoak, K., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press/Bradford Books.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797–2822.

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10,* 417–446.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 76,* 1723–1743.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development, 2,* 189–219.

Loewenstein, J., Thompson, L., & Gentner, D. (2003). Analogical learning in negotiation teams: Comparing cases promotes learning and transfer. *Academy of Management Learning and Education, 2,* 119–127.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215,* 1519–1520.

Opfer, J. E., & DeVries, J. M. (2008). Representational change and magnitude estimation: Why young children can make more accurate salary comparisons than adults. *Cognition, 108,* 843–849.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' *living things* concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49,* 301–332.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55,* 169–195.

Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79,* 790–806.

Opfer, J. E., Thompson, C. A., & Furlong, E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science, 13,* 761–771.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394.

Rattermann, M. J., & Gentner, D. (1998). More eveidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13*, 453–478.

Roberts, W. A. (2005). How do pigeons represent numbers? Studies of number scale bisection. *Behavioral Processes*, *69*, 33–43.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444.

Siegler, R. S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science*, *19*, 759–763.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science, Special Issue on Mathematical Cognition*, *11*, 655–661.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, & Education*, *3*, 143–150.

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effect of context on age and sex differences in magnitude estimation. *Journal of Experimental Child Psychology*, *101*, 20–51.

U.S. Department of Education (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Uttal, D., Schreiber, J. C., & DeLoache, J. S. (1995). Waiting to use a symbol: The effect of delay on children's use of models. *Child Development*, *66*, 1875–1889.

Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1787–1798

Differing Levels of Gender Salience in Preschool Classrooms: Effects on Children's Gender Attitudes and Intergroup Bias

Lacey J. Hilliard and Lynn S. Liben
The Pennsylvania State University

Developmental intergroup theory posits that when environments make social-group membership salient, children will be particularly likely to apply categorization processes to social groups, thereby increasing stereotypes and prejudices. To test the predicted impact of environmental gender salience, 3- to 5-year-old children ($N = 57$) completed gender attitude, intergroup bias, and personal preference measures at the beginning and end of a 2-week period during which teachers either did or did not make gender salient. Observations of peer play were also made at both times. After 2 weeks, children in the high- (but not low-) salience condition showed significantly increased gender stereotypes, less positive ratings of other-sex peers, and decreased play with other-sex peers. Children's own activity and occupational preferences, however, remained unaffected.

People vary along many dimensions such as age, race, political party, hair color, and favorite activities. Why do children use some dimensions rather than others to categorize people into social groups? And what (if any) consequences result from sorting people into those groups? These questions are at the center of a recent formulation of developmental intergroup theory (Bigler & Liben, 2006, 2007) that builds on several other intergroup approaches (Aboud, 2003; Cameron, Alvarez, Ruble, & Fuligni, 2001; Killen & McKown, 2005; Nesdale, 2004; Rutland, 1999; Rutland, Cameron, Milne, & McGeorge, 2005).

Developmental intergroup theory views the emergence of social stereotypes in children through constructivist lenses. Thus, as in other constructivist views of gender development (e.g., gender schema theory; see Martin & Halverson, 1981), children are thought to be active agents in processing social stimuli. Much early research in the constructivist tradition addressed how gender schemata influence the way that children process material differently depending on whether the material is gender traditional versus nontraditional and on whether it is culturally linked to the child's own versus the other sex. Illustrative of the former is research showing that children tend to have difficulty in remembering nontraditional material either by distorting it so that it becomes consistent with gender stereotypes or by forgetting it entirely (e.g., Koblinsky, Cruse, & Sugawara, 1978; Liben & Signorella, 1980). Illustrative of the latter is research showing that children are more likely to attend to, remember, and enact material associated with their own sex (e.g., Grusec & Brinker, 1972; Martin, Eisenbud, & Rose, 1995; Signorella, Bigler, & Liben, 1997). In developmental intergroup theory, these constructive processes are examined for their role in forming and maintaining social-group categories that in turn become the basis for developing stereotypes and prejudices.

Bigler and Liben (2006, 2007) argued that the development and application of children's stereotyping and prejudice are influenced by the salience of groups within children's surroundings. They identified various factors that render groups more or less salient. One factor concerns the degree to

---

The data reported here were collected as part of the first author's master's thesis, conducted under the supervision of the second author. Portions of this work were presented in April 2008 at the Gender Development conference, San Francisco. We express our appreciation to the child-care directors, Janie Hilliard and Barbara Crosby, for their willingness to accommodate this project, and to the participating staff, teachers, children, and parents for their friendly cooperation. We are also grateful to the members of the Cognitive and Social Development Lab at Penn State for their many contributions to data coding and entry, to Rebecca Bigler for her insights during various phases of this research, and to the Faris endowment of Penn State's Department of Psychology for financial support.

Correspondence concerning this article should be addressed to Lacey J. Hilliard or Lynn S. Liben, Department of Psychology, Moore Building, The Pennsylvania State University, University Park, PA 16802. Electronic mail may be sent to hilliard@psu.edu or liben@psu.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0013

which group membership is visually or perceptually observable (e.g., skin color vs. political affiliation). Irrespective of its initial visibility, a particular quality may be made more visible by an environment that marks group membership in some way. For example, men and women may wear different clothing and hair styles that exaggerate their physical differences, members of different religious groups may wear different head coverings or religious symbols, and members of different political parties may don distinctive lapel buttons. In addition, groups may be made more salient through the use of different language and labels to refer to group members (Bigler, Jones, & Lobliner, 1997).

Also contributing to the impact of group divisions is the child's own drive to categorize: As a part of their constructive, self-motivated drive to understand their interactions with their social and physical environment, children attempt to categorize their early environment and then make sense of these categorizations (Waxman & Gelman, 1986; Waxman & Markow, 1998). Thus, when children recognize differences between groups and are made aware of those groups through implicit and explicit categorization and labeling, they are likely to use those bases to divide people into groups. They are then internally and externally driven to attach meaning to those categories. For example, if religion is made salient in a child's environment (e.g., via differential clothing and labeling), young children would be expected to be more vigilant than they otherwise might be in seeking to find what is associated with religion. Children may also treat categories as innate due to their tendency toward essentialism (Gelman, 2003). Thus, if the child's environment divides people by race, children would be expected to be more likely to infer that differences between races rest on fundamental differences between the groups.

Children also have a propensity toward an in-group and out-group bias (Abrams, Rutland, & Cameron, 2003; Killen, Margie, & Sinno, 2006), and thus once children have enacted their categorization processes, they tend to view their own group as superior and the "other" group as inferior. Intergroup approaches have been used to study and understand stereotypes and prejudices in children. Early work showed that once children are assigned to groups, they make positive attributions to their own in-group and negative attributions toward the out-group (Sherif, Harvey, White, Hood, & Sherif, 1961). Consistent with the core tenets of developmental intergroup theory, these intergroup processes are heightened when there is a physical difference between groups (e.g., race; see Aboud, 1988). The importance of the distinction between in- and out-group attitudes has received recent attention, specifically in a call for measures that do not assume a reciprocal association between the two (Brewer, 1999). Recent investigators have also questioned the yoked nature of in-group preference and out-group derogation and have suggested that the two constructs are not inextricably linked (Aboud, 2003; Cameron et al., 2001; Kowalski, 2003).

One important arena in which the processes identified in developmental intergroup theory may operate is gender. Indeed, over a quarter century ago, Bem (1983) offered an analysis that is consistent with the ideas presented in developmental intergroup theory in that she argued that society's emphasis on gender categorization led to a host of dysfunctional and unnecessary sex differences. Despite this critique, children continue to encounter experiences that increase the salience of gender as when teachers greet their pupils with "Good morning, boys and girls" or when schools require that boys and girls wear different-colored robes at commencement ceremonies.

There has been only a small body of empirical research bearing on the impact of different levels of gender salience. Studies have shown the effect of gender labels on attributing traits to individuals (see Stern & Karraker, 1989), but few experiments have evaluated the role of using gendered language in the environment or have studied whether in-group preferences and out-group biases emerge as predicted by intergroup theories. In one of the few relevant studies, Bigler (1995) found that when teachers used gender to organize their elementary school classes, children showed greater gender stereotyping than did children in classrooms that avoided the functional use of gender. In this study, younger children with less advanced classification skills were more susceptible to the effects of the manipulation than older children (i.e., children with less advanced classification skills developed more negative attitudes than peers with more advanced classification skills).

Although not applied specifically to gender, Patterson and Bigler (2006) have provided evidence that even younger children are also susceptible to developing in-group biases based on group divisions. Specifically, they used a novel group paradigm in which preschool children were randomly assigned to different-colored shirts. They found that when the teacher made functional use of color groups (e.g., lining children up for recess by shirt color), children developed significantly stronger in-group biases. However, it remains unknown if preschool children would be affected by gender labeling and organization in classroom environments. Research to address this issue is important not only because of the practical implications for preschool classroom structure but also because it concerns a category division—gender—that is already emphasized heavily in young children's lives. It is possible that in this larger societal context, preschool classroom manipulations would have little impact.

The evidence just discussed suggests that young children's gender attitudes are susceptible to environmental manipulations that make social-group categories highly salient. Research in the intergroup literature has largely focused on the consequences of sorting people into groups and how one's sense of identity toward a group influences intergroup relations (Tajfel & Turner, 1986). Investigators have shown that gender is prevalent as an in-group category and have reported children's gendered play patterns and own-sex preferences (Maccoby & Jacklin, 1987; Yee & Brown, 1992, 1994). Developmental intergroup theory draws on the fields of intergroup relations but provides a model to test *how* stereotypes are formed and exaggerated, specifically by manipulating the salience of a social-group category in children's environments. The current study expands on previous work of intergroup processes and novel group paradigms by incorporating a wide range of measures to assess children's intergroup preferences and peer play behavior in the domain of gender. More specifically, the purpose of the present study was to test whether increasing the salience of gender in preschool classrooms would lead children to develop more highly gender-stereotyped attitudes and whether there would be evidence of both increased in-group and out-group bias (i.e., favoring one's own gender group and avoiding the other-gender group).

To address these questions, children were studied over a 2-week period under one of two gender-salience conditions. In the low-salience condition, children continued to experience a preschool environment in which the teacher avoided making gender explicit. In the high-salience condition, children experienced a preschool environment in which the teacher highlighted gender by using gender-specific language (labeling groups and individuals with gender terms) and by using gender-based organization in the classroom (e.g., assigning boys and girls to separate bulletin boards, lining up boys and girls separately for recess). Based on developmental intergroup theory, we predicted that children in the low-salience condition would show no discernible changes in their gender stereotypes over the 2-week period of the study, whereas children in the high-salience group would show a significant increase in their gender stereotypes (i.e., a decrease in gender flexibility).

If the increased salience of gender in the classroom was indeed found to strengthen children's gender stereotypes, developmental intergroup theory would predict that the high-salience condition should also yield stronger in-group and out-group bias. To test this latter prediction, we assessed children's interest in playing with peers of their own sex (in-group) and with peers of the other sex (out-group). Specifically, both prior to and following the classroom intervention, children were asked to rate their interest in playing with each of the other children in their class, and their average ratings for own- versus other-sex peers were compared across the two times. As an additional posttest measure, children were asked to select which of three groups of unfamiliar peers they would like to play with—one with only boys, one with only girls, or one with both boys and girls. Of interest was whether children in the high-salience classrooms would express greater reluctance to play with other-sex children than would children in the low-salience classrooms. As a third way to evaluate children's interest in own- versus other-sex peers, we collected observational data on children's peer play. We predicted that children in the high-salience (but not low-salience) classrooms would show a significant decrease in play with other-sex peers over the 2-week period. Measures in the present study were designed so that in- and out-group biases could be assessed independently (i.e., children could show positive in-group attitudes without showing negative out-group attitudes, and vice versa). Researchers have suggested a primacy of in-group preference compared to out-group derogation (Aboud & Amato, 2001; Brewer, 1999). If children's preference for in-group members is strongly established by the early preschool years whereas negative attitudes toward out-group members develop somewhat later, then an experimental manipulation in the preschool years would be more likely to increase negative attitudes toward out-group members rather than influence already-established positive attitudes to in-group members.

The predictions described earlier concern children's gender attitudes about others and the consequences of those attitudes (i.e., in-group and out-group biases). Another aspect of gender development concerns the gendered nature of children's own interests, for example, whether children favor toys and educational or occupational pathways that are culturally viewed as masculine or feminine. Past research has suggested that children's gendered personal interests are generally resistant to short experimental interventions, even when those interventions have a significant impact on their beliefs about what others should do (e.g., Bigler, 1995; Bigler & Liben, 1990). To examine whether a similar pattern of results would hold here, we also collected pre- and posttest data on personal interests in culturally stereotyped activities and occupations.

Method

Participants and School Context

Participants were 57 children ranging in age from 3 years 1 month to 5 years 6 months (M = 4.7 years; SD = 0.65 years). Children were roughly evenly divided between two preschools, each enrolling approximately equal numbers of boys and girls (14 boys and 16 girls in one school; 14 boys and 13 girls in the other). The majority of participants were European American from middle-class backgrounds. Informed consent was obtained from the directors of each school and teachers in each classroom. All parents and children were informed about data collection (interviews and observations) by letter and given the opportunity to opt out of the research activities if desired. None of the parents or children declined to participate.

The two schools (four classrooms) were located in a mid-sized city in the Southwest and were similar along important dimensions (e.g., teacher education, class size, teacher–child ratio, populations served, policies implemented, and classroom arrangement). Despite similarities, however, there are always differences between any two schools and between any two classrooms within a given school and grade (e.g., different teachers, different peers). Thus, the design included both pre- and posttests on critical measures so that analyses could take into account any initial differences between children in the two conditions.

Classroom Conditions

Classrooms in the high-salience condition were designed to make gender salient by exposing children to frequent uses of gender classifications in their environment. Following procedures used by Bigler (1995), teachers were encouraged to make frequent use of gender through physical separation (e.g., lining children up by sex), classroom organization (e.g., posting separate boys' and girls' bulletin boards), and gender-specific language (e.g., "I need a girl to pass out the markers" and "Good morning boys and girls"). Although simple statements or requests from a teacher using gender might be interpreted by the child as implying competition, we specifically asked teachers to avoid using gender distinctions for rewards or competitive events. Thus, teachers were asked to limit their use of gender to labeling and classroom organization, and explicitly asked to treat groups equally and to avoid encouraging competition between boys and girls (e.g., we asked teachers to avoid comments such as: "The boys are being quieter so they can line up" or "Let's see who can clean up faster, boys or girls").

Teachers in the low-salience classrooms were given no instructions about changing their behaviors in any way. Importantly, the ongoing policy in both preschools was to avoid using gendered language and organization in the classroom, and thus, children in both conditions had a similar classroom history with the exception of the 2-week period of the study proper. Researchers observed the classrooms every few days over the 2 weeks to observe whether teachers were continuing to use gender (or not) as requested. Indeed, teachers in the high-salience classrooms followed instructions by using gendered language and separating their classroom activities according to gender, whereas teachers in the low-salience classrooms did not organize their classrooms by gender.

Assignment to condition was made at the level of the school rather than at the level of classroom. One reason for this decision was the practical need to meet the administrative needs of the schools, which required having consistent expectations for teachers. Additionally, it carried two design benefits. First, it ensured consistent behaviors among teachers who sometimes move between classrooms or interact with children outside their assigned classroom. Second, it avoided the problem of contamination effects across classrooms, a problem that has been noted in earlier school-based intergroup research (Patterson & Bigler, 2006).

Procedure Overview

Children were interviewed individually both before and after the classroom manipulation with measures described next. In addition and also as explained in more detail later, children were observed during their normal classroom play periods in sessions at the beginning and end of the classroom manipulation. Following the 2-week period and the administration of all measures, children participated in a debriefing and educational program administered by both teachers and experimenters. This debriefing program was included in an effort to counteract any possible increase in stereotyping and to help children understand prejudice and stereotypes. During the program, puppets, pictures, and posters were used to emphasize similarities and differences among people that were not based on gender. Although we would have liked to have administered delayed assessments of children's gender stereotypes to study long-term effects of the classroom condition and the success of debriefing, practical limitations on the time that the school could devote to the research precluded adding additional assessments to the study.

Gender Attitudes

Children's endorsements of cultural gender stereotypes were assessed before and after the intervention period by the activity and occupation subscales of the Preschool Occupation, Activity, and Trait–Attitude Measure (POAT–AM), developed as a part of a suite of sex-typing measures (see Liben & Bigler, 2002). In this measure, children are shown pictures of activities or occupations, and for each item asked if men/boys, women/girls, or both men and women "should" perform it. Children were asked to respond to 66 occupations and activities, including 22 culturally masculine (e.g., use tools, be a firefighter), 20 culturally feminine (e.g., play with dolls, be a dancer), and 24 neutral (e.g., fly a kite, be a writer) items by pointing to one of three cards to show who they believed should do the activity. The three cards showed schematic (silhouette) pictures of people, with one card showing a man and boy, another showing a woman and girl, and the third containing all of these figures. Scores for this measure are the number of "both men and women" responses. Thus, lower numbers of "both" responses indicate higher endorsement of cultural gender stereotypes and less gender flexibility.

Peer Play

Familiar-peer ratings. To assess children's interest in playing with own- versus other-sex peers at both pre- and posttest, children were shown photographs of each of their classmates. For each photograph, the child was asked, "How much would you like to play with [name of classmate]?" Children gave their responses by pointing to one of three faces that comprised a 3-point scale. The faces included negative, neutral, and happy expressions, and were explained as meaning they had no interest at all, some interest, or strong interest. Two scores were calculated for each participant, one averaging ratings of children of their own sex, and the other averaging ratings of children of the other sex.

Unfamiliar-peer preference. An additional measure of children's gender-based peer play preferences was administered at posttest only. In this task, children were shown three photographs of groups of unfamiliar peers, one showing all boys, one showing all girls, and the third showing both boys and girls. Each child was told, "These are children from another school," and asked, "If you were to go to that other school, which group of kids would you like to play with?"

Peer play observations. Children were observed during indoor free play periods through a one-way mirror in each classroom. Observations were conducted during free play periods to ensure that teachers were not assigning children to groups in any way. Based on procedures developed by Martin and Fabes (2001), raters observed children sequentially in 15-s units by following a randomly ordered list of children in the classroom. Once raters had observed each child on the list, they started again at the top of the list, until there were three observational periods for each target child on each of 4 days.

Two observers independently noted the number of own- and other-sex peers the child played with during each of the observational periods. There was high reliability between coders on the number of boys and girls with whom the child played (interclass correlations = .96 and .93, respectively). Data from the primary observer were used in the rare cases of disagreement. Two scores were calculated for each period (the beginning and the end of the study): One score was the number of own-sex children and the other score was the number of other-sex children with whom the child played, averaged over the six observations (three observations for each of 2 days at each time period). The pretest observations were calculated from 2 days of observations made just before and as the classroom manipulation was beginning, and the posttest observations were calculated from 2 days of observations made just as and after the classroom manipulation was ending.

Personal Interests

Children's own gender-linked interests were assessed with the activity and occupation subscales of the POAT–Personal Measure (POAT–PM; see Liben & Bigler, 2002). In the POAT–PM, children are asked to indicate the strength of their own interests in various occupations and activities by pointing to one of three faces described earlier for the peer rating measure. Following standard administration procedures, the POAT–PM was administered prior to the POAT–AM to minimize cross-measure contamination.

Instructions were explained with a gender-neutral item, "How much would you like to eat ice cream?" Once children understood the instructions, they were shown the pictures for the POAT–PM proper, and for each item they were asked, "How much would you like to ____?" Responses of *no interest* were scored as 1, *some interest* as 2, and *strong interest* as 3. The dependent measures were calculated separately for masculine and feminine items and were computed as the total number of points awarded divided by the total number of items in that category (i.e., masculine or feminine).

Results

Overview and Preliminary Analyses

Results are described in three major sections that present, first, data on children's gender-stereotyped attitudes; second, data related to children's interest in or observed play with peers of their own or the other sex; and, third, data on children's own interests in stereotypically masculine versus feminine activities and occupations.

Given the relatively narrow range of ages of participants, we had no hypotheses with respect to age differences within our sample. Nevertheless, we conducted preliminary analyses to check for possible age effects both with respect to the way that measures were functioning (e.g., reliability of the POAT scales) and with respect to the effects of the experimental manipulation. The measures were comparably successful for younger and older groups (ages 3.1–4.3 vs. 4.4–5.5, respectively) and effects did not vary with age either in main effects or in interactions. Therefore, results reported are based on the entire sample.

In a second preliminary analysis, we examined correlations between pairs of peer variables to evaluate whether our dependent measures were tapping redundant constructs. Correlations were uniformly low, suggesting that as intended, the various measures were tapping different aspects of gender-related cognitions and behaviors. Thus, results are presented separately for each measure below.

Gender Attitudes

The dependent variable used to assess gender stereotypes was the total number of "both" (i.e., gender-flexible) responses given to masculine and feminine items on the POAT–AM. Given that the POAT–AM contains 22 masculine and 20 feminine items, scores could range from 0 to 42, with higher scores indicating more flexibility (i.e., lower gender stereotyping). Scores were analyzed with a repeated measures analysis of variance (ANOVA) in which between-subjects variables were condition (low vs. high salience) and participant sex (boys vs. girls), and the within-subjects variable was time (pre- vs. posttest). When appropriate, t tests were used as follow-up tests. Differences reported as significant were $p < .05$ or better.

Analyses showed a significant interaction between condition and time, $F(1, 53) = 17.85$, $p < .001$, $\eta_p^2 = .25$. As shown in Figure 1, although in the low-salience condition there was no significant change in gender-stereotyped attitudes from pre- to posttest, M (SD) = 14.46 (7.14) versus

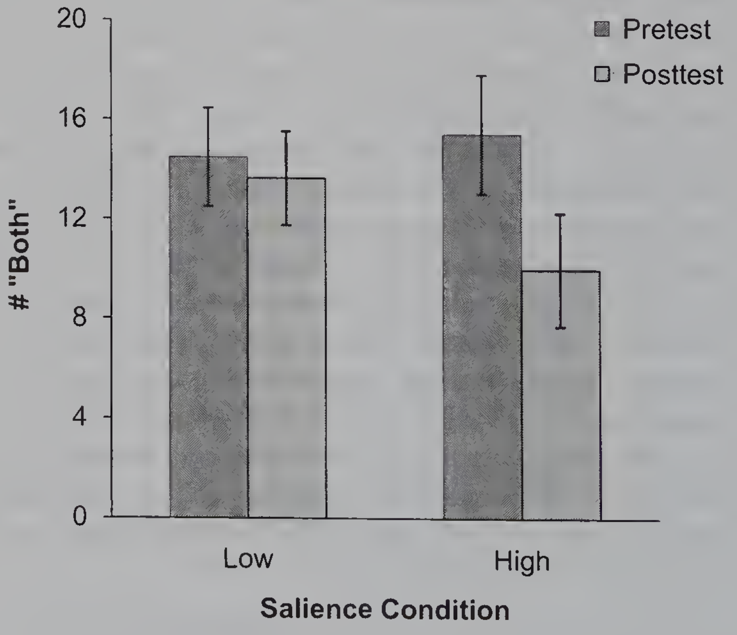


Figure 1. Mean numbers (and standard errors) of "both" (flexible) responses on the POAT–AM gender attitude measure by condition and time.
Note. Higher scores indicate greater gender flexibility or lower gender stereotypes (maximum score = 42).

13.61 (8.13), in the high-salience condition there was a significant drop in the number of flexible "both" responses from pre- to posttest, *M* (*SD*) = 15.37 (6.99) versus 9.94 (6.18), *d* = 0.12, indicating an increase in gender stereotypes. Also significant was the interaction between time and participant sex, *F*(1, 53) = 5.75, *p* = .020, η_p^2 = .09. The drop in the number of "both" responses over time was not significant in boys, 14.07 (7.00) versus 12.00 (8.00), but was significant in girls, 18.89 (6.23) versus 13.14 (8.15), *d* = 0.70.

Peer Play

Familiar-peer ratings. To examine children's ratings of interest in playing with familiar peers, the average ratings of boy and girl classmates served as the dependent measure in a repeated measures ANOVA. Between-subjects variables were condition (low vs. high salience) and participant sex (boys vs. girls), and within-subjects variables were time (pre- vs. posttest) and peer type (own vs. other sex).

Analyses revealed a significant three-way interaction among condition, time, and peer type, *F*(1, 48) = 8.46, *p* = .005, η_p^2 = .14, depicted graphically in Figure 2. In the low-salience condition, there was no significant change in other-sex peer ratings between pre- and posttest, *M* (*SD*) = 2.21 (0.47) versus 2.36 (0.37), whereas in the high-salience condition, there was a significant decrease in other-sex ratings over time, *M* (*SD*) = 2.09 (0.59) versus 1.48 (0.36), *d* = 1.03. Subsumed by the three-way interaction were two-way interactions between time and condition, *F*(1, 48) = 14.26, *p* < .001, η_p^2 = .22, and between peer type and condition, *F*(1, 48) = 16.39, *p* < .001, η_p^2 = .26.

Unfamiliar-peer preference. The dependent variable for assessing interest in playing with unfamiliar peers was the type of group selected, that is, the participant's selection of an own-sex group (i.e., the boy-only group for boys or the girl-only group for girls) versus an other-sex group (i.e., either the other- or the mixed-sex group). In both the high- and low-salience groups, only 1% of children chose to play with the other-sex only groups (i.e., those with all boys or all girls). For the purpose of analyses, other- and mixed-sex groups were pooled given that the important question concerned the child's willingness to play with the other sex. Data showed that 37% of children in the low-salience condition chose to play with a group that included children of the other sex compared to only 13% in the high-salience condition, a marginal effect of condition, $\chi^2(1) = 3.71$, *p* = .054.

Peer play observations. To assess children's observed peer play, we calculated average numbers of own- and other-sex peers per observational time unit (15 s). Play with only other-sex and mixed-sex children was collapsed for statistical purposes to cover play with any type of play with other-sex peers. A repeated measures ANOVA was conducted in which between-subjects variables were condition (low- vs. high-salience) and participant sex (boys vs. girls), and within-subjects variables were time (beginning vs. end of 2-week period) and type of peer (own- vs. other-sex).

Analyses revealed a significant three-way interaction among condition, time, and type of peer, *F*(1, 47) = 25.46, *p* < .001, η_p^2 = .35, shown in Figure 3. As may be seen in the figure, children in the low-salience condition showed no significant change in play with other-sex peers between the beginning and end of the 2-week period, *M* (*SD*) = 0.64 (0.30) versus 0.74 (0.39), whereas children in the high-salience condition showed a significant decrease in the amount of play with other-sex peers over time *M* (*SD*) = 0.92 (0.41) versus 0.26 (0.19), *d* = 1.3. Subsumed by the three-way interaction were a marginal two-way interaction between condition and time, *F*(1, 47) = 3.90,

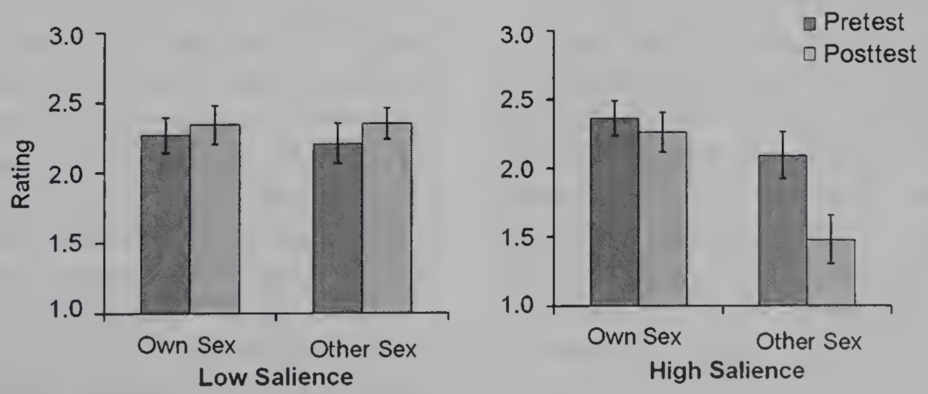


Figure 2. Mean ratings (and standard errors) of children's interest in playing with familiar peers by condition, peer type, and time (most positive rating = 3).

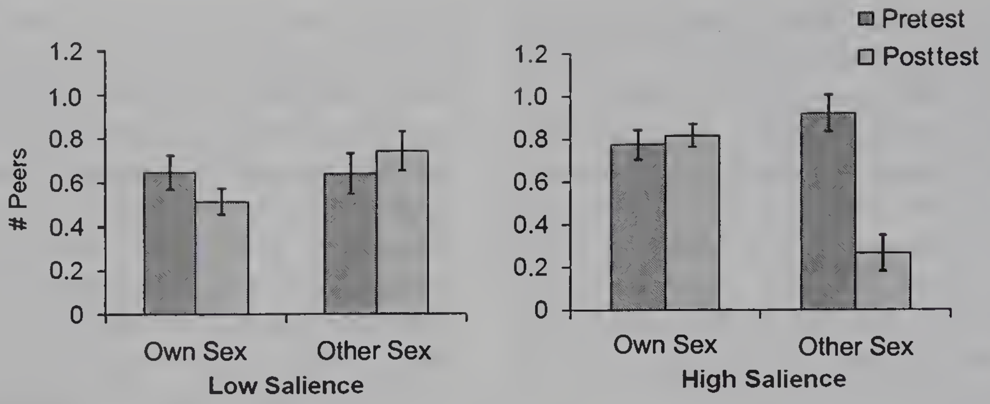


Figure 3. Number of peers (and standard errors) played with, on average, during 15-s observational units by condition, peer type, and time.

$p = .054$, $\eta_p^2 = .08$, and a two-way interaction between time and type of peer, $F(1, 47) = 6.04$, $p = .018$, $\eta_p^2 = .11$.

Personal Interests

As explained earlier, personal interest in culturally gender-stereotyped activities and occupations was assessed by the POAT–PM. Average interest scores (on a 3-point scale) were calculated separately for masculine and feminine items and then analyzed with an ANOVA. Between-subjects variables were condition (low vs. high salience) and participant sex (boys vs. girls), and within-subjects variables were time (pre- vs. posttest) and item type (masculine vs. feminine).

Analyses revealed a significant interaction between participant sex and item type, $F(1, 48) = 205.60$, $p < .001$, $\eta_p^2 = .81$. As expected, boys preferred masculine items more than feminine items, 2.50 (0.24) versus 1.86 (0.28), $d = 2.2$, whereas girls preferred feminine items more than masculine items, 2.48 (0.21) versus 2.11 (0.35), $d = 1.05$. No other main effects or interactions were significant.

Discussion

The major purpose of this research was to test predictions derived from developmental intergroup theory concerning the impact of environmental qualities on the development of stereotypes and prejudices. Specifically, we tested the predictions that making gender more salient in the classroom would lead first, to stronger gender stereotypes and second, to stronger intergroup biases in preschool children. Data provided compelling support for both predictions.

With respect to gender stereotypes, data from the POAT–AM measure showed that children in classrooms in which gender was made salient expressed significantly more highly stereotyped attitudes at posttest than they had at pretest. A comparable change did not occur among children in the low-salience classrooms, indicating that increased stereotyping in the high-salience condition cannot be attributed merely to retesting effects or to the passage of time.

With respect to in-group versus out-group biases, data from several directions converge on the conclusion that children who experienced high-salience classrooms developed increased rejection or avoidance of the out-group (i.e., the other sex). One line of evidence for this conclusion comes from the familiar-peer ratings. Children in the high- (but not low-) salience condition showed a pre- to posttest decrement in their ratings of how much they would like to play with their other-sex classmates. A second line of evidence comes from the unfamiliar-peer preference measure given at posttest. This measure revealed that children in the high-salience classrooms were marginally ($p = .054$) less likely than children in the low-salience classrooms to say they would like to play with groups of unfamiliar peers that included other-sex children. Finally, a third line of evidence derives from observational data on children's free play in the classroom which revealed that over the 2-week period of the study, children in the high-salience condition showed a significant decrease in their play with other-sex peers whereas children in the low-salience condition showed no such decline.

It is interesting that despite the fact that the measures used could have revealed increased favoritism toward the in-group as well as increased rejection of the out-group, our data showed evidence for the latter only. That is, even in the high-salience group, there were no significant increments over time in children's ratings of own-sex peers or in their actual play with own-sex peers. In earlier work on race-based social groups, Cameron et al. (2001) found that preschool-aged children held positive in-group bias toward their own racial group, but did not demonstrate negative out-group evaluations, suggesting a primacy for in-group bias. Aboud (2003) found that in-group positivity and out-group negativity were reciprocally correlated for children with little personal experience with out-group members (i.e., from a racially homogenous school), but in-group positivity and out-group negativity were not reciprocal for children from racially heterogeneous school. In the present study—with gender-defined social groups and coeducational classes—children had experienced extensive daily interactions with other-sex peers, thus offering no reason to expect a reciprocal association. However, it is also risky to assume equivalent processes for groups defined by race and gender, and thus it would be important to continue to explore both in-group and out-group processes and the way that they may be linked in the domain of gender.

Before leaving the discussion of the data relevant to in-group and out-group bias, it is important to point out that when children were tested under the conditions of their normal classroom environment (i.e., at both pre- and posttest for the low-salience condition; at pretest for the high-salience condition), the data showed no evidence of rating own-sex children higher than other-sex children (Figure 2) or of playing more with own-sex children than other-sex children (Figure 3). Given that other investigators have found evidence of own-sex preferences in preschool settings (e.g., Martin & Fabes, 2001), one possible interpretation of the current data is that preschools that assiduously avoid gender-based divisions or language in their classrooms indeed prevent (or postpone) in-group biases observed in other settings. The data may also reflect the egalitarian values of parents who enrolled their children in these schools.

Although more research is thus needed to learn precisely when and how in-group and out-group gender biases evolve, the data from the current study already demonstrate the role of social-group salience on bias, and bolster arguments made by others (Aboud, 2003; Allport, 1954; Brewer, 1999; Cameron et al., 2001) about the importance of assessing in-group and out-group biases with distinct measures.

Taken together, the findings on children's gender stereotypes and group bias extend earlier work with respect to both age and scope of effects of social-group salience. Earlier work (Bigler, 1995) had demonstrated that increasing gender salience had led to increased stereotypes in older (elementary school) children; the current data demonstrate that these forces operate even among younger children who have less advanced categorization skills. Recent research (Patterson & Bigler, 2006) had shown that highly visible, novel groups that had been defined by colored shirts led to stereotypes and group biases in preschool children. The current data demonstrate that similar effects result from increasing the salience of gender, even though it is already a social category well known to preschoolers (e.g., Ruble, Martin, & Berenbaum, 2006; Weinraub et al., 1984).

As in any study, many important questions remain for future research. One issue concerns how long the observed salience effects persist. As noted earlier, practical constraints prevented the inclusion of delayed posttests in this study. However, having now established immediate and dramatic effects of gender salience in the classroom, it becomes important for future research to address the degree to which such effects are long lasting, and to identify the factors that influence their longevity.

Another question concerns the associations among the outcome variables. We had selected a range of measures addressed to constructs hypothesized to be affected by the gender-salience manipulation. One measure (POAT–AM) was selected to assess the degree to which the child endorsed cultural gender-stereotyped attitudes about what others should do. Two measures (peer ratings and, at posttest only, the novel peer measure) were selected to assess children's cognitions about their interest in playing with others of their own and the other sex. One measure (play observation) was selected to study children's actual behavior with their own- and other-sex peers. All measures for which there were pre- and posttest data showed a significant impact of the manipulated classroom condition, and the posttest only novel peer measure showed a marginal effect of group. Thus, the data from the current study allow us to conclude that as hypothesized, increasing gender salience in the classroom affects a range of children's gender-related beliefs and behaviors. Remaining for later research with much larger samples is the intriguing and important question about how the various components of gender beliefs and behaviors are connected, in some cases, perhaps causally. For example, future research might examine whether a rise in gender stereotyping decreases children's willingness to play with children of the other sex, or whether there is an inverse pathway, or if perhaps each outcome is instead controlled by some other factor.

A clear implication of the findings already in hand is that schools should make it as unacceptable to use gender-specific language and divisions (e.g., ''Good morning boys and girls'') as it is to use race-specific language and divisions (e.g., ''Good morning Black children and White children''). There are ample anecdotal data that the former continues to occur in our educational system. For example, as this article was being written, the second author received a note from a former student reporting that she had worked over the summer in a preschool center ''and they actually split the children up for free play time, saying girls you can play this and boys you can play that, i.e., dolls and action figures'' (D. Russell, personal communication, October 7, 2008).

What also remains for future research is exploration of the cognitive mechanisms by which increased gender salience leads to greater gender stereotypes and out-group biases. As proposed by Bigler and Liben (2006, 2007), developmental intergroup theory posits that one causal mechanism for stereotyping lies in children's active attempts to assign meaning and correlates to social categories that are made salient in the environment. To test this position, it would be necessary to monitor children's cognitive processing (e.g., measure children's attention to implicit and explicit correlates of gender) over time under low- versus high-salience conditions.

Another potential causal mechanism involves the peer environment. Previous work has shown that peer groups provide a powerful context to socialize gender-linked play styles and preferences (Fabes, Martin, & Hanish, 2003; Maccoby & Jacklin, 1987). In a longitudinal study in preschool classrooms, Martin and Fabes (2001) found that children who were engaged in high levels of own-sex play at the beginning of the study showed more gender-typed play at the end of the study than did children who were engaged in low levels of own-sex play. They noted that over time, the play styles of each gender group became increasingly distinct.

Several aspects of our work showed that the high-gender-salience condition led children to reject other-sex peers, thereby increasing gender segregation. First, anecdotes reported by teachers in the high-salience condition suggest that children were quick to pick up on gender categorization and organize themselves by gender. One teacher, for example, noted that on the second day of the manipulation, without prompting from her, children established separate boys' and girls' snack tables. Second, the formal observational data on peer play showed that children in the high-salience condition became less likely to play with children of the other sex over time. To the degree that increased gender salience in the classroom led children to play less with members of the out-group, children would have even less occasion to be socialized into the play styles and preferences typical of the other sex, in turn exaggerating differences between gender groups. It is likely, given previous findings and our current study, that early gender segregation leads to two distinct gendered peer cultures, which, exacerbated over time, establish a script for later relationship styles and affect life decisions (Leaper, 1994).

In short, anecdotal data on children's classroom behaviors and systematic data on children's peer play suggest that much of the effect of the teachers' greater use of gender in the high-salience groups may have had its impact by inspiring children themselves to create a highly gendered environment. Another important direction for future research is therefore to observe, in detail, how the classroom environment (e.g., children's communication, segregation, and role playing) evolves under low- versus high-salience conditions.

Finally, the data from the current research add to the literature concerning the strong and persistent nature of children's own gendered interests. As expected, responses to the POAT–PM showed that children held sex-linked preferences, with boys preferring masculine items more than feminine items and girls preferring feminine items more than masculine items. There was no modification of this pattern as a function of gender-salience condition. The finding is consistent with earlier research (e.g., Bigler, 1995; Bigler & Liben, 1990) showing that brief experimental interventions leave children's own gendered interests and preferences unchanged even in the face of significant changes in children's beliefs about what others should do. Such findings suggest that modifying children's gender stereotypes will not be sufficient to expand children's educational and occupational choices (e.g., encouraging more girls and women to pursue science-related hobbies, coursework, and careers) and that more directed intervention strategies are needed. It is plausible that the 2-week intervention was not long enough to affect children's own interests and that over a longer period of time, children would be likely to internalize the gender stereotypes and show more gender-linked interests.

In conclusion, the findings from the present research have important implications for both theory and practice. At the theoretical level, data provide support for a core prediction of developmental intergroup theory (Bigler & Liben, 2006, 2007), which posits that social-group salience plays an important role in the formation of stereotypes and biases. At the practical level, data demonstrate that there is a pervasive, powerful, and remarkably quick effect of making gender salient in the classroom. The current study highlights the impact in one area of a child's environment (the classroom), but one would also expect a strong influence of the attention to social-group categories in the home (e.g., different chores for male and female members of the household) and in the broader culture (e.g., different professional sports leagues) as well. Although there is much to learn about the mechanisms by which increasing the salience of group membership operates, the current research demonstrates the importance of minimizing attention to social-group divisions in young children's classroom environments, even for social-group divisions, like gender, that are already visible and salient in the society at large.

References

Aboud, F. E. (1988). *Children and prejudice*. New York: Blackwell.

Aboud, F. E. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology, 39*, 48–60.

Aboud, F. E., & Amato, M. (2001). Developmental and socialization influences on intergroup bias. In G. J. O. Fletcher & M. S. Clark (Series Eds.) & R. Brown & S. Gaertner (Vol. Eds.), *Blackwell handbook in social psychology: Vol. 4. Intergroup processes* (pp. 65–85). Oxford, UK: Blackwell.

Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development, 74*, 1840–1856.

Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs: Journal of Women in Culture and Society, 8*, 598–616.

Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development, 66*, 1072–1087.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68*, 530–543.

Bigler, R. S., & Liben, L. S. (1990). The role of attitudes and intervention in gender-schematic processing. *Child Development, 61*, 1440–1452.

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). San Diego, CA: Elsevier.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotypes and prejudice. *Current Directions in Psychological Science, 16*, 162–171.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429–444.

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about ingroups and outgroups: Reconceptualizing research on prejudice. *Personality and Social Psychology Review, 5*, 118–128.

Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development, 74*, 921–932.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press.

Grusec, J. E., & Brinker, D. B. (1972). Reinforcement for imitation as a social learning determinant with implications for sex-role development. *Journal of Personality and Social Psychology, 21*, 149–158.

Killen, M., Margie, N. G., & Sinno, S. (2006). Morality in the context of intergroup relations. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 155–184). Mahwah, NJ: Erlbaum.

Killen, M., & McKown, C. (2005). How integrative approaches to intergroup attitudes advance the field. *Journal of Applied Developmental Psychology, 26*, 616–622.

Koblinsky, S. A., Cruse, D. F., & Sugawara, A. I. (1978). Sex-role stereotypes and children's memory for story content. *Child Development, 49*, 452–458.

Kowalski, K. (2003). The emergence of ethnic and racial attitudes in preschool-aged children. *Journal of Social Psychology, 143*, 677–690.

Leaper, C. (1994). Exploring the consequences of gender segregation on social relationships. *New Directions for Child Development, 65*, 67–86.

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67*(2, Serial No. 269).

Liben, L. S., & Signorella, M. L. (1980). Gender-related schemata and constructive memory in children. *Child Development, 51*, 11–18.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 195–216). New York: Academic Press.

Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's gender-based reasoning about toys. *Child Development, 66*, 1453–1471.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*, 431–446.

Martin, C. L., & Halverson, C. F. (1981). A schematic-processing model of sex typing and stereotyping in children. *Child Development, 52*, 1119–1134.

Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 219–245). New York: Psychology Press.

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development, 77*, 847–860.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon & R. M. Lerner (Series Ed.) & N. Eisenberg (Vol. Ed.) *Handbook of child psychology* (6th ed., Vol. 3, pp. 858–932). New York: Wiley.

Rutland, A. (1999). The development of national prejudice, in-group favouritism and self-stereotypes in British children. *British Journal of Social Psychology, 38*, 55–70.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development, 76*, 451–466.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University Book Exchange.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1997). A meta-analysis of children's memories for own-sex and other-sex information. *Journal of Applied Developmental Psychology, 18*, 429–445.

Stern, M., & Karraker, K. H. (1989). Sex stereotyping of infants: A review of gender labeling studies. *Sex Roles, 20*, 501–522.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 7–24). Chicago: Nelson-Hall.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1*, 139–156.

Waxman, S. R., & Markow, D. B. (1998). Object properties and object kind: Twenty-one-month old infants' extension of novel adjectives. *Child Development, 69*, 1313–1329.

Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, R., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-typed toy preferences, and family characteristics. *Child Development, 55*, 1493–1503.

Yee, M. D., & Brown, R. (1992). Self-evaluations and intergroup attitudes in children aged three to nine. *Child Development, 63*, 619–629.

Yee, M. D., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33*, 183–196.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1799–1813

Race Salience and Essentialist Thinking in Racial Stereotype Development

Kristin Pauker
Stanford University

Nalini Ambady
Tufts University

Evan P. Apfelbaum
Northwestern University

The authors explored the emergence and antecedents of racial stereotyping in 89 children ages 3–10 years. Children completed a number of matching and sorting tasks, including a measure designed to assess their knowledge and application of both positive and negative in-group and out-group stereotypes. Results indicate that children start to apply stereotypes to the out-group starting around 6 years of age. Controlling for a number of factors, 2 predictors contributed significantly toward uniquely explaining the use of these stereotypes: race salience (i.e., seeing and organizing by race) and essentialist thinking (i.e., believing that race cannot change). These results provide insight into how and when real-world interventions aimed at altering the acquisition of racial stereotypes may be implemented.

Young children possess a number of abilities that may serve as the foundation for learning racial knowledge. Three-month-old infants can discriminate perceptually between different racial groups (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2005), preferring to look at faces that belong to a familiar racial group. By 3–4 years of age, children can sort people by race (Aboud, 1988; Nesdale, 2001), and soon after, those in the ethnic majority group (e.g., European American children in the United States) show signs of implicit and explicit in-group preferences (Baron & Banaji, 2006; Cameron, Alvarez, Ruble, & Fuligni, 2001; Nesdale, 2004). Recent work has explored several variables underlying biased group attitudes, such as the role of categorization (Bigler, Jones, & Lobliner, 1997; Patterson & Bigler, 2006) and status (Bigler, Brown, & Markell, 2001; Nesdale & Flesser, 2001). However, despite notable research exploring children's racial attitudes, relatively little work has investigated the emergence and antecedents of children's racial stereotyping.

Much of past research exploring children's racial stereotyping has muddled the distinction between attitudes and stereotypes. Social psychologists have defined stereotypes as *cognitive* structures composed of consensual knowledge, beliefs, and expectations about social groups that may result in *both* positive and negative associations for a single specific group (Hamilton & Trolier, 1986). Stereotypes are thus distinct from racial attitudes, which reflect affective evaluations or preferences, where one group is consistently considered more positively and another more negatively. Measures conceived with the intent to measure stereotyping, such as "trait stereotyping," often measure a construct more akin to attitudes (see Ruble et al., 2004). We attempted to clarify the literature on children's stereotyping by investigating how children learn both positive and negative stereotypes about their in-group and out-group.

Determining *when* children begin to use racial stereotypes is essential for understanding their developmental trajectory. Research has established that racial stereotyping occurs in early adolescents (Blake & Dennis, 1943; Maykovich, 1972; Rowley, Kurtz-Costes, Mistry, & Feagans, 2007); however,

---

We are extremely grateful to all the children, parents, teachers, and administrators who helped make this project possible. We would also like to thank Archana Asundi, Jillian Russo, Megan Oswald, Monica Malowney, Robyn Yano, and Sara Kaplan for their assistance with data collection and coding. This research was supported by a Jacob K. Javits Fellowship to Kristin Pauker and National Institute of Health Award (R01 MH070833-02) granted to Nalini Ambady.

Correspondence concerning this article should be addressed to Kristin Pauker, Department of Psychology, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305. Electronic mail may be sent to kpauker@stanford.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0014

limited research exists with younger children, especially those below 10 years of age. Thus, we set out to systematically explore children's use of stereotypes pertaining to European American, African American, and Asian American individuals beginning at 3 years—the youngest age at which children have been reported to exhibit the ability to sort by race (Nesdale, 2001)—a necessary precursor to racial stereotyping. By exploring stereotyping among children 3–10 years of age, we were able to encompass an age range wide enough both to examine the developmental acquisition and eventual application of such stereotypes.

Equally essential to understanding racial stereotype development is determining *what* factors facilitate the emergence of these stereotypes. The few studies that have explored children's racial stereotypes, as defined before, have either concentrated on the behavioral consequences of stereotype awareness (Ambady, Shih, Kim, & Pittinsky, 2001; Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; McKown & Weinstein, 2003) or the content of stereotypes pertaining to a few specific groups (Bar-Tal, 1996; Bigler, Averhart, & Liben, 2003; Corenblum, 2003). Far less attention has been devoted to the social and cognitive variables associated with racial stereotype development.

Predictors of Stereotyping

Our understanding of social development can greatly benefit from integrating developmental and social psychological approaches (Bigler & Liben, 2007; Olson & Dweck, 2008). In one such integrative approach, Bigler and Liben (2007) propose that three core processes contribute to the formation of social stereotypes and prejudice: (a) establishing the psychological salience of different person attributes, (b) categorizing individuals by salient dimensions, and (c) developing sets of traits, behaviors, and emotions that are associated with particular salient social groups. Additionally, they hypothesize key factors—perceptual discriminability, proportional group size, and explicit and implicit use—shape the operation of psychological salience, the first of these core processes. Once a dimension becomes salient, they suggest that essentialism, in-group bias, and explicit and implicit attributions mold the development of stereotyping and prejudice pertaining to that dimension. We focused on two components this model suggests should affect racial stereotyping: race salience and essentialist thinking.

Race Salience

Racial stereotyping must be preceded by an understanding that race is a meaningful basis for social categorization. Merely noticing a perceptual difference by which one *can* sort people does not mean this difference will be perceived as a meaningful, psychologically salient dimension. Labeling and both explicit and implicit use of a category increase its salience (Bigler et al., 1997; Patterson & Bigler, 2006). For example, children may notice that shirt color is a dimension by which they can sort people, but they may not use it until they are engaged in a capture-the-flag game where shirt color denotes team members. Subsequently, children start to organize their world based on salient dimensions and infer that any observed differences in behavior must be caused by inherent group differences (Bigler & Liben, 2007). Returning to our example, if the team in blue beats the team in red in three consecutive games, children may infer those on the blue team are inherently athletic. Thus, after a category has become psychologically salient to a child, essentialist thinking may foster stereotype development.

Essentialist Thinking

Essentialist thinking—the tendency to think of categories as immutable and informative—is grounded in the belief that certain categories have important underlying essences that define their nature and properties (Gelman, 2003; Medin & Ortony, 1989). A number of studies with adults have linked essentialist beliefs to stereotyping (e.g., Plaks, Stroessner, Dweck, & Sherman, 2001; Yzerbyt, Corneille, & Estrada, 2001; Yzerbyt, Rocher, & Schadron, 1997). Essentialist beliefs even predict racial stereotyping in adults to a greater extent than, and independent of, a number of established individual difference measures known to be associated with stereotyping (Bastian & Haslam, 2006; Levy, Stroessner, & Dweck, 1998).

While studies show children essentialize both animal (e.g., Gelman & Wellman, 1991) and human (e.g., race and gender; Hirschfeld, 1995; Taylor, 1996) categories, few have explored the social consequences of essentialist thinking in children. In one of the few studies to address this issue, Levy and Dweck (1999) demonstrated that children who believed that people's personal characteristics were fixed and immutable stereotyped novel group members to a greater degree. Immutability is one of the central components of essentialist thinking, and thus it is likely to be an important factor in the emergence of children's racial stereotypes.

Children naturally achieve a state where they view categories through such an essentialist lens when they attain racial constancy—a social-cognitive transition defined by viewing racial group membership as immutable. Once children believe that their own and others' racial group membership is permanent, they may consequently search out explanations for perceived similarities within or differences between racial groups (Cameron et al., 2001). Rutland, Cameron, Bennett, and Ferrell (2005) recently demonstrated that preschool children's level of racial constancy was significantly related to their racial bias scores (see also Semaj, 1980). While racial constancy has been related to racial bias, it remains an open question as to whether racial constancy also impacts racial stereotyping. Thus, we examined a naturally occurring change in essentialist thinking—the achievement of racial constancy—as a predictor of children's racial stereotyping.

Considerable variability has been reported with respect to *when* children achieve racial constancy—some work suggests racial constancy does not emerge until at least 7 or 8 years (Aboud, 1988; Aboud & Ruble, 1987; Hughes, 1997) and other work reports children exhibiting such capacities as early as preschool (Hirschfeld, 1995). These studies may in fact be measuring different components of similar constructs. Children may have nascent theories about race in preschool but may not have an explicit understanding of a coherent essentialist theory about race until slightly later. Examining children's social reasoning often provides valuable insight into children's understanding of social concepts not available through simple forced-choice measures (e.g., Killen & Stangor, 2001; Ruble et al., 2007; Taylor, Rhodes, & Gelman, 2009). Drawing on this framework, we probed children's reasoning regarding racial constancy to gauge age-related changes in essentialist thinking and reasoning.

The Present Research

The present investigation focused on two primary aims. First, we examined the emergence of children's racial stereotypes. Second, we explored potential antecedents to their development.

Since children start to demonstrate ethnic and racial awareness as young as 3 or 4 years of age and reliably report a number of stereotypes by early adolescence, we chose to investigate racial stereotyping in children 3–10 years of age. Our measure included both positive and negative stereotypes associated with three racial groups: European American, African American, and Asian American, which we recoded into in-group relevant and out-group relevant stereotypes. We predicted most children would demonstrate racial stereotyping around 6 years, since (a) previous work has demonstrated children's awareness of race-related occupational stereotypes at 6 years of age (Bigler et al., 2003) and (b) children at this age either have reached or will soon reach several important social-cognitive milestones, including racial constancy (Ruble et al., 2004). Accordingly, we first explored in-group and out-group stereotypes across four age groups: a very young group where we expected little or no stereotyping (preschoolers: 3–5 years), an older group where we expected prevalent stereotyping (second to fourth graders: 7–10 years), and two narrower groups centered around 6 years (kindergarteners: 5–6 years, first graders: 6–7 years) where we expected stereotyping to emerge.

The timing of stereotype acquisition may also depend on a child's own group membership. We expected that by virtue of being the numerical minority, minority group members who are more psychologically salient and distinctive (McGuire & McGuire, 1988) will more often be the targets of racial stereotyping than proportionally larger (racial majority) groups (Bigler & Liben, 2007). Thus, we predicted that European American, racial majority group children should exhibit more stereotyping with age, but that they would primarily use out-group opposed to in-group stereotypes.

Next, we examined several factors predicted to contribute to the emergence of racial stereotyping. To control for children's level of cognitive development, we included a classic sorting task often employed as a measure of classification skills. This provided us with a measure of cognitive skill and an opportunity to assess children's spontaneous use of race as a sorting dimension. Children's use of race as a sorting dimension and their mention of race in a subsequent task in which they described an array of faces, served as measures of race salience. Finally, children completed a racial constancy task, which included a free-response question in which we coded their use of essentialist reasoning. While we expected both age and classification skill would be related to racial stereotyping, we predicted that two primary factors would contribute uniquely to the emergence of racial stereotype awareness in children: race salience and essentialist thinking in racial constancy.

Method

Participants and Design

A total of 89 children were recruited from the local community through mail solicitations to parents and from four suburban public elementary schools that serve middle- and upper-middle-income families outside Boston, Massachusetts. The return rate of initial letters sent out to the local community was 40%, and 90% of those parents agreed to participate. In the elementary schools, among consents sent home with students in participating classes, 75% gave written permission. These children's schools were predominantly European American (76%), with some Asian American (11%), African American (4%), and Latino American (9%) children; their neighborhoods also had a similar composition. The sample consisted of four age groups: preschoolers (3–5 years; $M = 3.73$ years, $SD = 0.37$; $n = 21$ [9 females]), kindergarteners (5–6 years; $M = 5.29$ years, $SD = 0.50$; $n = 36$ [14 females]), first graders (6–7 years; $M = 6.71$ years, $SD = 0.25$; $n = 16$ [9 females]), and second to fourth graders (7–10 years; $M = 8.97$, $SD = 1.32$; $n = 16$ [7 females]). Participants were mostly European American (89.9%), with a small representation of minority participants (4.5% Asian American, 5.6% Multiracial).

Measures and Procedure

Parents were informed of the study via letters sent home by school administrators or by phone for those children recruited from the community. Upon receipt of parental consent, individual children who provided verbal assent participated in a quiet location, separate from other children in the elementary school or separate from their parents. The experimenter led the participant through four tasks, assessing racial stereotyping, classification skill, race salience, and essentialist reasoning in racial constancy. To avoid contamination from tasks that were obviously about race (e.g., racial constancy) and to enable spontaneous assessment of children's use of race as a categorization dimension in earlier tasks (e.g., classification task), these tasks were always presented in the same order.

Racial Stereotyping

Forty-eight high-resolution photos of male children were arranged into 24 pairs. We used only male photographs for the racial stereotyping task because several of the stereotypes used are more strongly associated with males (e.g., Maddox & Gray, 2002; Niemann, Jennings, Rozelle, Baxter, & Sullivan, 1994). Photographs were taken from the Internet and an existing laboratory database and were pretested for perceived attractiveness, age, race, and emotionality. Adults ages 18–22 ($N = 30$) and children ages 6–10 ($N = 15$) rated the photos. Adults used a 7-point scale and children used a 5-point scale to rate attractiveness. For the remaining items, raters indicated their free response. Photos receiving extreme ratings of attractiveness or age were dropped, as were photos for which racial categorization was not reliably accurate. Each of the 48 final selected photos was cropped below the shoulders, gray-scaled, and pasted onto a uniform background before being placed into a photo album. The experimenter led the child through the photo album while narrating 24 brief behavioral episodes.

For each episode, participants were asked to select the child in the depicted pair that was most likely to have demonstrated a target behavior (see Ambady et al., 2001). Nine target episodes described behaviors that typify prevalent *negative* stereotypes about three racial groups: African Americans (i.e., stealing from others, acting aggressively, underperforming academically), Asian Americans (i.e., being shy, acting submissively, retaining foreign customs), and European Americans (i.e., bragging, ordering others around, excluding others). Nine target episodes described *positive* stereotypes about the same three racial groups: African Americans (i.e., being musical, playing basketball well, being rhythmic), Asian Americans (i.e., being hardworking, playing violin skillfully, excelling in math), and European Americans (i.e., being wealthy, receiving preferential treatment, acting as a leader). Younger children (3- to 4-year-olds) only heard a subset of the stereotypical episodes (the first two positive and negative episodes listed for each racial group, totaling six negative and six positive episodes), to increase the likelihood that their concentration would last through the whole task. Six nonstereotypical fillers described race-neutral behaviors (i.e., liking animals, playing outdoors, eating chocolate, listening to stories, making things with play-doh, playing on the swings). Stereotypical behaviors were chosen based on older children's (above 8 years) and early adolescents' open-ended racial stereotypes (Blake & Dennis, 1943; Maykovich, 1972; Zeligs, 1947) and commonly reported adult racial stereotypes (Lin, Kwan, Cheung, & Fiske, 2005; Madon et al., 2001; Niemann et al., 1994). The stereotypicality or neutrality of the episodes was validated through pretesting with adults.

We specifically chose to focus on racial stereotypes relevant to a school setting (e.g., academics, extracurricular activities, peer interactions), because not only may racial stereotypes limit children's potential in a number of academic and extracurricular domains, they may also exacerbate intergroup relations through enforcing self-segregation into "acceptable" stereotype-consistent domains (Kao, 2000).

For stereotypical episodes, we always displayed one photo of a child belonging to the racial group (African American, Asian American, or European American) targeted by the stereotype. The second photo was chosen from among the other racial groups with the stipulation that the photo chosen did not belong to the same racial group as the target photo. The racial group of the child not associated with the stereotype was randomized for each episode and counterbalanced across participants. This ensured that participants who belonged to one of the racial groups in the pair were not always making a decision that involved an in-group/out-group choice. For example, if a participant was European American, an African American stereotypical episode could include an in-group (European American) target and an out-group (African American) target or two out-group (Asian American and African American) targets. For filler episodes, the racial composition of the pairs was randomized. We counterbalanced the order in which episodes were presented and whether the target photo appeared on the left or the right. Additionally, we randomized the pairings of photo sets and episodes.

Participants practiced with two randomly selected filler episodes before moving on to the remaining 22 behavioral episodes. For example, in a behavioral episode used to gauge a positive African American stereotype, a participant would be told, "One of these children is really good at basketball. Everyone wants him on his team because they will win. Which of these boys do you think he probably is?" Participants were instructed to point to the child most likely to have demonstrated the behavior, and their responses were recorded. An answer was coded as stereotypical if a child picked the target in a stereotypical episode. We averaged participants' stereotyping scores across out-group episodes and in-group episodes. Thus, if a participant was European American, out-group stereotyping was composed of Asian American and African American episodes and in-group stereotyping was composed of European American episodes.

Classification Task

Using procedures reported in Bigler (1995), the classification task was composed of 16 photographs of people cropped at the waist who varied systematically on several dimensions: race (African American, European American), sex, age (children, adults), and facial expression (serious, laughing). The photographed individuals also differed with respect to a variety of other nonsystematic dimensions (e.g., clothing, posture).

The experimenter spread all the photos on a table and asked the participant to sort them into two groups such that, "people who go together are in the same pile." If they were able to successfully complete a sort, they were asked to explain their sort. After the initial sort, the experimenter asked the participant if there was an additional way to sort the photos. This continued until the participant could no longer think of dimensions upon which to sort the photos.

Children who did not complete a successful sort were given a score of 0. Children who were able to sort the stimuli according to one dimension but were not able to re-sort the stimuli in a new way were given a score of 1. Children who were able to complete two such sorts were given a score of 2 (see Bigler & Liben, 1993). When scoring the sorts, coders checked that each child's explanations for their sorts matched the piles they created. For this and all subsequent coding tasks, two experimenters independently coded the responses. Overall, interrater agreement was high (Cohen's $\kappa = .98$), and disagreements were resolved via discussion.

Race Salience

Race salience was gauged by two measures: sorting by race in the preceding classification task and racial justification. First, those who spontaneously sorted by race in any of their sorts in the preceding classification task were given a score of 1 and those who did not sort by race at all were given a score of 0 for the sort by race variable.

Second, we measured children's use of racial justification in a second task where they described sets of photos. In this task, two arrays of six same-gender photographs, one all-White and one racially diverse, were presented to participants, counterbalanced across trials. Participants were asked whether the people in the photographs were the same or different and why they were the same or different (see Semaj, 1976). We coded for participants' use of race or skin color as a reason for being different in the racially diverse set or as a reason for being the same in the all-White set. Children who acknowledged racial difference in either of the sets were given a score of 1 and those who did not were given a score of 0 for the racial justification variable.

The sort by race and racial justification variables were not combined into one index. Although they both indicate race salience, the first measure (sort by race) more clearly represents functional use of race, whereas the second measure (racial justification) more clearly represents noticing and describing a perceptual difference in the array. The stimuli used in the sort by race measure varied along many systematic dimensions. Thus, children's choice to sort by race meant that they preferentially used that dimension or perceived that dimension as more salient than the others, reflecting its functional use. In contrast, the stimuli used in the racial justification measure appeared similar on all dimensions, except for race, and we specifically asked children what made the pictures the same or different. Therefore, the stimuli intentionally highlighted racial differences and restricted available alternative descriptions. As a result, children's choice to report that they noticed racial differences did not involve making any trade-offs between other potential dimensions. Because of this conceptual difference, we kept the two measures separate since others have found that functional use of categories exacerbates intergroup bias compared to just noticing perceptual differences (Bigler, 1995; Bigler et al., 1997), and thus the two measures of race salience could differentially influence racial stereotyping.

Essentialist Thinking in Racial Constancy

Racial constancy was assessed both for judgments of the self and others, since we reasoned that such judgments might differentially predict a child's focus on in-group or out-group stereotypes. We modified previous tasks used to assess racial or gender constancy (Hirschfeld, 1995; Ruble et al., 2007; Semaj, 1980), targeting three core components of children's understanding of racial constancy: identification, stability, and consistency (Ruble et al., 2004).

Judgments of the self. Identification (i.e., correct identification of one's own category membership) was assessed using the question, "Are you a White big boy/girl or a Black big boy/girl?" Most participants were European American and received a White/Black comparison pair; however, if they were Asian American or multiracial, we asked their parent or teacher for the most appropriate term to use and used Black as the alternate category. Two questions were used to assess stability (i.e., understanding race remains constant over time): "When you were a baby, were you a White baby or a Black baby?" and "When you grow up to be a man/woman will you be a White man/woman or a Black man/woman?" Lastly, consistency (i.e., understanding race is consistent across superficial transformations) was assessed by the initial question, "If you really wanted to change your skin color to be a Black man/woman could you?" and a follow-up question to reveal the reasoning underlying their initial answer (e.g., "How would you do that?" or "Why not?").

Judgments of others. We assessed stability and consistency for judgments of others using photographs matched to the participant's gender. Again, we used two indicators to assess stability: First, participants saw either a photograph of a White or Black child above one White and one Black adult, and were asked, "When this child grows up, will they look more like this adult or that adult?" Second, participants saw a photograph of a White or Black adult above one White and one Black child, and were asked, "When this adult was little, did they look more like this [White] child or this [Black] child?" To assess consistency, participants were shown a picture of a White child and were asked the initial question, "If this child really wanted to be Black and change his/her skin color could he/she do that?" Children were then asked a follow-up question to reveal the reasoning underlying their initial answer (e.g., "How would he/she change?" or "Why can't he/she change?").

Racial constancy and essentialist reasoning. Children responses to the items relating to identification and stability were coded as correct (1) or incorrect (0). For the consistency item, we also took into account children's reasoning in coding their answer (see Ruble et al., 2007). First, children's underlying reasoning was coded for content. Children only received a (1) for the consistency item if the initial question was answered correctly *and* if the follow-up question indicated true constancy by utilizing essentialist reasoning (e.g., "He can't change. You're born one way and you can't change after that"). Finally, the scores were added together to form a composite score, with higher scores indicating a greater sense of racial constancy for the self or others.

In sum, we employed two types of broad measures: one type that gauged use of in-group and out-group stereotypes and a second type including an assortment of variables that gauged capacities we hypothesized to be associated with the emergence of such stereotype usage. The predictor variables included factors we wanted to control for (i.e., classification skill) and two main predictors of interest: race salience (i.e., sort by race, racial justification) and essentialist thinking (i.e., composite racial constancy score for self, composite racial constancy score for others, essentialist reasoning used in the racial constancy measures).

Results

Overview of Analyses

Our primary outcome of interest was participants' use of racial stereotypes, averaged across in-group and out-group relevant episodes. Recall, participants chose which child was most likely to exhibit a behavior, and each target episode was coded as in-group stereotype or out-group stereotype relevant. First, we explored potential age group differences in children's use of both positive and negative stereotypes associated with in-groups and out-groups using a three-way analysis of variance (ANOVA). Next, we employed hierarchical multiple regressions to examine the possibility that increased understanding of race salience and essentialist thinking in racial constancy may explain unique variance in children's racial stereotype development not accounted for by other variables, such as age and classification skill.

Stereotyping

To explore age group differences in positive and negative in-group and out-group stereotyping, we submitted the stereotyping scores averaged across items to a 2 (group: in-group, out-group) × 2 (valence: positive, negative) × 4 (age group: 3–5 years, 5–6 years, 6–7 years, 7–10 years) ANOVA, with repeated measures on the first two factors (see Table 1 for all means). In a first pass, we included sex and race of the participant in the analyses, but no differences were obtained as a function of sex, race, or any of their interactions so analyses collapsed across these variables. Notably, we also found no effects of valence or interactions with valence, *F*s < 0.53, *p*s > .47; thus, valence was not considered further. Only the main effect of age group, $F(3, 85) = 6.95$, $p < .001$, $\eta^2 = .20$, and the Group × Age Group interaction, $F(3, 85) = 3.14$, $p = .03$, $\eta^2 = .10$, were reliable. As displayed in Figure 1, children's stereotyping (both positive and negative) increased with age (main effect of age group), and children's in-group and out-group stereotyping emerged at different ages (Group × Age Group interaction). Follow-up simple-effects tests used to clarify this interaction indicated that age group differences were reliable for out-group, $F(3, 85) = 12.38$, $p < .0001$, $\eta^2 = .30$, but not in-group stereotypes, $F(3, 85) = 0.994$, $p = .40$, $\eta^2 = .03$. Bonferroni multiple comparisons used to explore the reliable age group differences in out-group stereotyping revealed that the two older groups exhibited more out-group stereotyping overall than the two younger groups. Specifically, two oldest groups ($p > .99$) and the two youngest groups ($p = .12$) did not differ significantly in their out-group stereotyping; however, the 6- to 7-year-olds ($M = 0.60$, $SD = 0.12$) exhibited more out-group stereotyping than 3- to 5-year-olds ($M = 0.38$, $SD = 0.18$, $p < .001$) and marginally more out-group stereotyping than 5- to 6-year-olds ($M = 0.49$, $SD = 0.15$, $p = .08$). The 7- to 10-year-olds ($M = 0.67$, $SD = 0.13$) exhibited more out-group stereotyping than either of the younger groups (*p*s < .001).

Table 1
In-Group and Out-Group Positive and Negative Stereotyping by Age Group

Stereotype valence (year-olds)	In-group stereotyping		Out-group stereotyping	
	Positive	Negative	Positive	Negative
3–5	0.55 (0.41)	0.59 (0.34)	0.35 (0.25)	0.41 (0.25)
5–6	0.53 (0.32)	0.57 (0.25)	0.47 (0.22)	0.50 (0.20)
6–7	0.55 (0.25)	0.49 (0.24)	0.60 (0.15)	0.60 (0.23)
7–10	0.60 (0.39)	0.69 (0.35)	0.67 (0.23)	0.67 (0.15)

Note. Values are given as *M* (*SD*).

Next, we ran one-sample *t* tests to determine whether racial stereotyping occurred at above-chance levels (50%). Despite a lack of relative differences across age groups for in-group stereotyping, children did use in-group stereotypes significantly above chance in the oldest age group (7–10 years; $M = 0.65$, $SD = 0.32$), $t(15) = 1.85$, $p = .04$, $r = .43$. Consistent with age group differences for out-group stereotyping, children in the two oldest age groups (6–7 years and 7–10 years; means reported before) utilized out-group stereotypes above chance, $t(15) = 3.38$, $p = .002$, $r = .66$ and $t(15) = 5.17$ $p < .0001$, $r = .80$, respectively. Given the strong differences in out-group stereo-

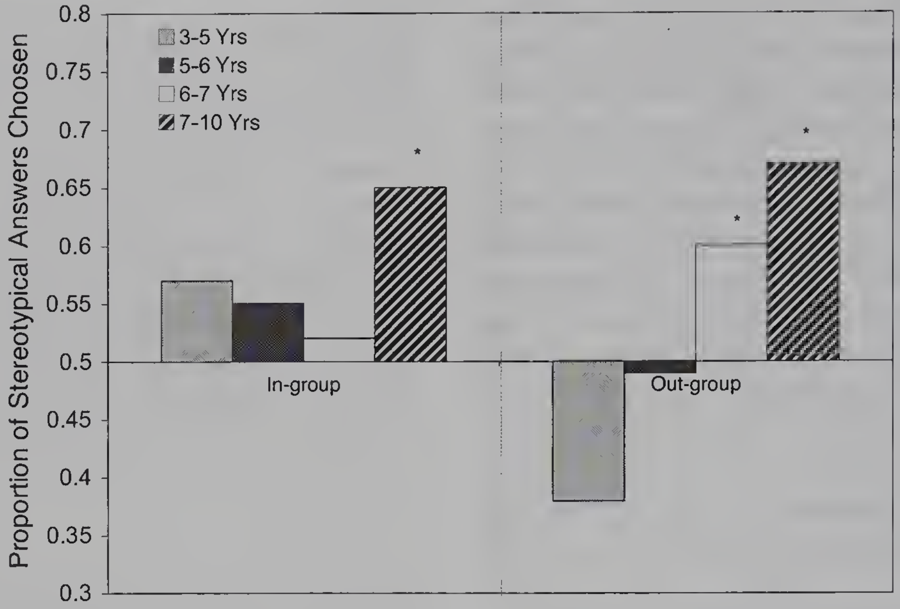


Figure 1. In-group and out-group stereotyping across age group. *Note.* Asterisk indicates a significant difference from chance.

typing and the lack of relative age group differences in in-group stereotyping, we sought to explore predictors of the emergence of out-group stereotyping. We were particularly interested in whether race salience and essentialist thinking in racial constancy uniquely predicted the emergence of out-group stereotypes.

Predictors of Out-Group Stereotyping

Age Group Differences in Predictors

Although not a central focus of this article, Table 2 provides means for age group differences across all of our predictors. One-way ANOVAs revealed children improved with age on all of our predictors ($Fs > 7.8$, $ps < .001$). At least half of the children exhibited perfect classification skills, sorted by race, and used a racial justification by 5–6 years of age. Racial constancy for self and others emerged slightly later, with the majority of children achieving a perfect score around 6–7 years of age.

Hierarchical Regression

Because of the significant positive relations between racial stereotyping and many of the predictors (Table 3), we performed multiple hierarchical regressions to determine the extent to which each predictor contributed unique variance in explaining patterns of out-group stereotyping. All regressions and correlations used age as a continuous variable. Seventy-four subjects were included in this analysis, since 12 subjects did not complete racial constancy measures due to time constraints, and an additional 3 subjects had incomplete answers for the racial justification task.

As displayed in Table 4, in Steps 1–3, we entered a number of factors for which we wanted to control in the regression: sex (0 = male, 1 = female), ethnic background (European American, Asian American, or Multiracial), children's age (in years), and classification skill (0, 1, 2). In Step 4, we entered our first predictors of interest: factors that indicated race salience (i.e., sorting by race and using a racial justification) coded as 0 = no and 1 = yes. In Step 5, we entered our final predictors of interest: factors that utilized essentialist thinking (i.e., children's racial constancy for self score and children's racial constancy for others score, respectively).

The model was significant at the second step, $F(4, 69) = 4.97$, $R^2 = .22$, $p = .001$, and remained significant across all subsequent steps through the final step, $F(9, 64) = 6.67$, $R^2 = .48$, $p < .001$. Once all variables were taken into account in the final model (see Table 4), none of the background variables (i.e., sex or race of the participant) significantly influenced out-group stereotyping. Although both age and classification skill contributed significantly to the model in earlier steps, once our predictors of interest were entered they no longer accounted for significant and unique proportions of variance in children's out-group stereotyping. As predicted, the factors that remained significant in the final model were those related to race salience and essentialist thinking: sorting by race ($\beta = .34$, $p = .003$), racial constancy for the self ($\beta = -.37$, $p = .005$), and racial constancy for others ($\beta = .49$, $p < .0001$). Note, however, that only one of our

Table 2
Age Group Differences in Predictor Variables

	Classification skills	Race salience	Racial justification	Racial constancy	
				Self	Others
3- to 5-year-olds					
Mean	0.86_{a} (0.85)	0.14_{a} (0.36)	0.00_{a} (0.00)	1.10_{a} (0.97)	1.24_{a} (0.83)
% max score	28.6	—	—	0.00	4.80
5- to 6-year-olds					
Mean	1.72_{b} (0.51)	0.50_{b} (0.51)	0.53_{b} (0.51)	2.69_{b} (1.24)	1.89_{b} (0.79)
% max score	75.0	—	—	33.3	19.4
6- to 7-year-olds					
Mean	2.00_{b} (0.00)	0.81_{b} (0.40)	0.56_{b} (0.51)	3.38_{b} (0.89)	2.63_{c} (0.62)
% max score	100	—	—	56.3	68.8
7- to 10-year-olds					
Mean	2.00_{b} (0.00)	0.80_{b} (0.45)	0.80_{b} (0.45)	3.20_{b} (1.30)	$2.40_{b,c}$ (0.89)
% max score	100	—	—	60.0	60.0
Range	0–2	0–1	0–1	0–4	0–3

Note. Values are given as *M* (*SD*). Means in the same column that do not share subscript letters differ at *p* < .05.

Table 3
Correlations Among Variables

Variable	1	2	3	4	5	6	7	8	9
1. Sex (female = 1)	—								
2. Age	.07	—							
3. Classification skills	.10	.62***	—						
4. Sort by race	−.03	.42***	.42***	—					
5. Racial justification	−.01	.43***	.25*	.42***	—				
6. Racial constancy for self	.16	.65***	.46***	.28*	.27*	—			
7. Racial constancy for others	.19	.61***	.57***	.17	.26*	.52***	—		
8. Stereotyping	.14	.26*	.43***	.35**	.24*	.19	.45***	—	
9. In-group stereotyping	−.04	.20	.06	−.05	.00	.08	−.04	.24*	—
10. Out-group stereotyping	.23*	.57***	.40***	.38***	.29*	.15	.48***	.61***	−.06

*p < .05. **p < .01. ***p < .001.

measures of race salience remained significant in the final model, namely, sorting by race. Racial justification was not a significant predictor in the final model, which highlights perhaps the importance of functional use of race (i.e., using it to organize the world) opposed to merely noticing perceptual differences. In contrast, both predictors of racial constancy contributed significantly toward explaining out-group stereotyping. Specifically, racial constancy for others was associated with increased out-group stereotyping, whereas racial constancy for the self was associated with decreased out-group stereotyping. This suggests that it is essentialist thinking specific to out-groups—not essentialist thinking construed more broadly—that predicts out-group stereotyping.

Essentialist Reasoning

We delved further into children's reasoning in the racial constancy measures by coding their responses regarding whether they or another person could change their race. Responses were coded into one of four categories: no answer provided, superficial reasoning (e.g., "paint my skin," "change my clothes"), essentialist reasoning (e.g., "you can't change your skin," "White mommies have White babies"), or idiosyncratic reasoning (e.g., "grow," "don't want to change"). We then employed chi-square analyses to compare the frequency of each type of reasoning across age groups (3–5 years, 5–6 years, 6–7 years, 7–10 years). Sixty-eight subjects were included in this analysis, since 6

Table 4
Final Model of Hierarchical Regression Analyses for Out-Group Stereotyping

	B	$SE\ B$	β	R^2	F change in R^2
Step 1: Background variables				.06	1.58
Sex	.05	.03	.13		
Race dummy: Asian American	−.20	.14	−.14		
Race dummy: Multiracial	−.09	.07	−.11		
Step 2: Age				.22	14.22***
Age	.02	.02	.11		
Step 3: Classification skill				.27	3.99*
Classification skill	.02	.03	.08		
Step 4: Race salience				.33	3.29*
Sort by race	.12	.04	.34**		
Racial justification	.03	.04	.08		
Step 5: Racial constancy				.48	9.36***
Self	−.05	.02	−.37**		
Others	.09	.03	.49***		

Note. Coefficients for each variable reflect values at Step 5.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

additional subjects completed the study without being asked why racial identity could or could not change.

Both judgments of the self, $\chi^2(9, n = 68) = 21.68$, $p = .01$, $\varphi = .56$, and others, $\chi^2(9, n = 68) = 31.04$, $p < .001$, $\varphi = .68$, changed across age groups. As displayed in Table 5, the patterns of reasoning across age groups was similar for judgments of both the self and others, except in the two younger groups. Reasoning among 3- to 5-year-olds was split between not providing a rationale for their answers reasoning ($M = 46.7\%$) and superficial reasoning ($M = 33.3\%$) for judgments about the self. These children generally did not provide a rationale for judgments about others ($M = 65.0\%$). Five- to 6-year-olds were split between not providing a rationale ($M = 20.0\%$), superficial reasoning ($M = 28.6\%$), and essentialist reasoning ($M = 37.1\%$) for judgments about the self. For judgments about others, however, these children used essentialist reasoning the most ($M = 46.7\%$) and superficial reasoning the second most ($M = 26.7\%$). For the two older age groups, children primarily used essentialist reasoning across both types of judgments (6- to 7-year-olds: $M_{\text{self}} = 78.6\%$, $M_{\text{other}} = 85.7\%$; 7- to 10-year-olds: $M_{\text{self}} = 75.0\%$, $M_{\text{other}} = 75.0\%$). However, only essentialist reasoning about *others* predicted out-group stereotyping, $r = .26$, $p = .03$. Essentialist reasoning about the *self* did not predict in-group or out-group stereotyping, $rs = -.07, .15$; both $ps > .21$.

Discussion

This research investigated the development and antecedents of children's racial stereotypes. These initial displays of stereotyping, around 6 years of age, were mainly restricted to out-group stereotypes but, importantly, involved the use of both positive and negative stereotypes. Our results provide empirical evidence that two critical factors—race salience and essentialist thinking—uniquely contribute to children's out-group stereotyping. Such a finding dovetails with recent theoretical work outlining the bases of racial stereotyping (Bigler & Liben, 2007). Consistent with a social-motivational model (e.g., Nesdale, 1999), neither age nor a measure of cognitive skill independently predicted out-group stereotyping. Children only oriented themselves toward out-group stereotypes once race became a functionally salient dimension and they developed a sense of the out-group as an essential category.

Stimulus salience has been shown to play a pivotal role in how individuals process social category information (Taylor, 1981). As race becomes a salient dimension to children, they may begin to organize information around racial categories and consequently form social stereotypes. Note that it was *not* children's mere mention of race as a descriptor (i.e., their use of a racial justification in describing a set of faces) that predicted the acquisition of out-group stereotypes but rather children's spontaneous use of race as a sorting dimension. In other words, children's functional *use* of race as an important organizing dimension in their world may facilitate racial stereotyping above and beyond children merely *noticing* perceptual racial differences (see Bigler et al., 1997; Bigler et al., 2001). As predicted, European American, racial majority children, who tend to have a less salient racial identity than racial minority children in the United States (McGuire & McGuire, 1988; Phinney, 1992), learned more stereotypes about out-group members before stereotypes about their own group. Thus, others' but not their own racial group membership was a distinctive dimension they noticed and used to formulate stereotypes.

Finally, children's understanding of racial constancy played a significant role in their use of out-group stereotypes, providing support for several theories that have proposed such a relation should exist (e.g., Aboud, 1988; Kohlberg, 1966). Racial constancy for *others* predicted increases in out-group stereotyping and racial constancy for the *self* predicted decreases in out-group stereotyping. One

Table 5

Percentage of Children Who Report Different Types of Reasoning to Explain Why or Why They Could Not Change to a Different Race

Type of reason (year-olds)	Racial constancy for self				Racial constancy for others			
	None	Superficial	Essentialist	Idiosyncratic	None	Superficial	Essentialist	Idiosyncratic
3–5	46.7	33.3	6.7	13.3	65.0	10.0	15.0	10.0
5–6	20.0	28.6	37.1	14.3	13.3	26.7	46.7	13.3
6–7	7.1	0.0	78.6	14.3	7.1	7.1	85.7	0.0
7–10	0.0	25.0	75.0	0.0	0.0	25.0	75.0	0.0

difference between the two tasks that could explain this difference in predictive power is that one relied primarily on images (racial constancy for others), whereas the other required knowledge of verbal labels (racial constancy for the self). While a plausible explanation, both tasks required knowledge of a verbal label to pass the task. Instead, we believe that this difference may be accounted for by the ease with which children essentialize the out-group versus the in-group and the effect that essentialist thinking applied to the out-group or in-group has on intergroup perceptions. Future research should certainly attempt to disentangle these factors and how they affect the emergence of racial stereotyping.

Although not a central focus of this study, children reliably demonstrated racial constancy, specifically for others, around 6 years of age. While this finding is inconsistent with Hirschfeld's (1995) work, it supports other studies on racial constancy (Rutland et al., 2005). Our measure also examined children's explanatory constructs, which may tap a more nuanced understanding of children's essentialist thinking, whereas Hirschfeld's method may tap a nascent biological theory. Importantly, this examination of children's explanatory constructs revealed that they steadily utilized more essentialist reasoning with increases in age, and that essentialist reasoning, specifically about others, predicted increases in out-group stereotyping. Thus, as children start to essentialize out-groups, they may also become aware of stereotypes associated with these groups.

The mechanisms involved in children's acquisition of racial stereotypes are likely to be complex and influenced by myriad factors. While race salience and essentialist thinking may provide the scaffolding for racial stereotype acquisition, the content of these stereotypes is drawn from cultural knowledge in the child's social milieu. For example, explicit information presented in books, TV, or educational curricula, and implicit or explicit information communicated by family, teachers, or friends may provide specific stereotype content (Bar-Tal, 1997). Additionally, implicit patterns of information available in the environment, such as segregation in schools or stratification of certain occupations in society may also furnish stereotype content (Bigler, Arthur, Hughes, & Patterson, 2008; Bigler & Liben, 2007). A number of additional factors, besides race salience and essentialist thinking are likely to be implicated in racial stereotyping as well (Bigler & Liben, 2007). For example, children's acquisition of racial stereotypes may be intertwined with their conceptions of status. Studies with adults have argued that more powerful groups (namely, the racial majority) attribute more stereotypes to those lower in the hierarchy to justify and legitimize their power (Yzerbyt et al., 1997). Young children are aware of status differentials (Bigler et al., 2003; Nesdale & Flesser, 2001) and, like adults, may try to rationalize observed inequities and status differences in society, serving to reinforce stereotypes of lower status groups.

While race salience may be a natural developmental acquisition in an environment where race is functionally important, racial stereotyping may not be an inevitable consequence of racial categorization or cognitive development. Certain social-contextual factors encourage more or less essentialist thinking about race, which provide scaffolding for children to make errors in judgment (i.e., stereotype) based on race. Future research should explore the types of contexts that minimize essentialist thinking about race and the types of educational strategies that curb the impact of pervasive explicit and implicit racial associations available in the environment. Moreover, this study suggests an implied underlying model of how age, cognitive skills, racial constancy, race salience, and essentialism work in concert to influence racial stereotyping. Future research should test this model explicitly using path analysis or structural equation modeling.

Relation Between Racial Stereotyping and Attitudes

While we have concentrated solely on racial stereotyping, racial attitudes surely constitute a sizable section of a child's racial knowledge toolbox. Children exhibit a reliable in-group bias early on, some as young as 3, well before the emergence of racial stereotyping documented in the present study (e.g., Nesdale, 2001). While many have interpreted this in-group bias as evidence that children possess negative racial attitudes, others have questioned the validity of this claim, arguing these findings reflect in-group preferences rather than true out-group derogation (e.g., Cameron et al., 2001; Pfeifer et al., 2007). Thus, while young children can label and organize their world by race early on—skills conducive to in-group preferences—they might not be able to symbolize an underlying concept and its associated attributes until they form a coherent essentialist theory about race.

The recognition of race as a psychologically salient dimension, early social preferences, and biological lay theories that all presumably develop in the preschool years provides a solid foundation upon which this developing system of social cognition—including both racial attitudes and stereotypes—can be built. Similar to racial stereotyping, a number of motivational factors may contribute to racial preferences becoming full-fledged racial attitudes, such as an explicit norm of prejudice in the environment or tension over limited resources (Nesdale, 2004). Thus, while children may have the foundation for racial attitudes and stereotypes early on, a host of motivational factors promote their consolidation into underlying concepts that guide behavior.

Limitations

Our sample included mainly racial majority (i.e., European American) children. The findings and models did not change when the sample of Asian American and multiracial children were removed from analyses. The one exception was that the comparison of in-group stereotyping to chance was no longer significant in the oldest age group—consistent with our conclusion that European American, majority children are primarily learning out-group stereotypes.

It will be important for future research to explore these processes with racial minority children. Since minorities in the United States have racial identities that are often perpetually salient and distinctive (McGuire & McGuire, 1988; Phinney, 1992), they may be more likely to learn both in-group relevant and out-group relevant stereotypes at an earlier age. Crucially, research should also consider how these processes work specifically with multiracial children whose existence challenges binary views of race. Racial constancy should not necessarily be thought of as the optimal "end-state." Multiracial children's racial identification often changes with context or time (Hitlin, Brown, & Elder, 2006), and they may develop more flexible theories about race that can mitigate the negative effects of stereotyping.

The diversity of children's surrounding environment is another important factor to consider. Our sample of children attended schools and lived in neighborhoods primarily composed of European Americans. Both school and neighborhood diversity may shape racial stereotyping in important ways, through changing children's individual cognitions (e.g., level of racial awareness), the function of stereotypes in that particular environment, and exposure to explicit and implicit racial associations. Future research should explore factors that predict the emergence of racial stereotypes in minority and multiracial populations and how structural aspects of the environment can shape the emergence of stereotyping. Additionally, although our study deliberately focused on a wide age range, future research should home in on factors that magnify race salience and exacerbate essentialist thinking in 3- to 6-year-olds.

On Preventing Stereotypes

Given that once stereotypes are formed they are highly resistant to change, especially in adults, and can affect critical real-world outcomes, such as children's academic performance (Ambady et al., 2001; McKown & Weinstein, 2003), our results lend important insight into age-appropriate interventions that may result in meaningful change. While change should be possible, since past studies have demonstrated that children's racial attitudes and preferences are malleable (e.g., Cameron, Rutland, & Brown, 2007; Katz & Zalk, 1978), less research has provided evidence for substantial change in children's racial stereotyping (but see Killen & Stangor, 2001). This may be due to a number of social-cognitive mechanisms that actually support stereotype maintenance in children (Bigler & Liben, 1993; Corenblum, 2003).

This research suggests that two potential factors play a key role in stereotyping racial out-groups: race salience and race-specific essentialist thinking.

Mitigating these factors should precipitate positive change. Interventions aiming to reduce race salience may be difficult to implement without making large-scale changes in a child's environment; however, recent studies with adults have demonstrated the ability to manipulate essentialist thinking about groups (e.g., Levy et al., 1998; Plaks et al., 2001). Thus, interventions aimed altering racial stereotypes may best be accomplished by directly manipulating children's pattern of thinking about racial groups. If children learn that race is not necessarily an immutable category and that even if the physicality of skin color is often immutable, it does not necessarily convey stable information about psychological attributes, perhaps they will be less susceptible to stereotyping others and to the adverse effects of stereotyping themselves.

References

Aboud, F. E. (1988). *Children and prejudice*. Oxford, UK: Blackwell.

Aboud, F. E., & Ruble, D. N. (1987) Identity constancy in children: Developmental processes and implications. In T. M. Honess & K. M. Yardley (Eds.), *Self and identity: Individual change and development* (pp. 95–107). London: Routledge and Kegan Paul.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385–390.

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology*, *44*, 1513–1518.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood: The case of "the Arab" concept formation, stereotype, and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations*, *20*, 341–370.

Bar-Tal, D. (1997). Formation and change of ethnic and national stereotypes: An integrative model. *International Journal of Intercultural Relations*, *21*, 491–523.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228–235.

Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, *66*, 1072–1087.

Bigler, R. S., Arthur, A. E., Hughes, J. M., & Patterson, M. M. (2008). The politics of race and gender: Children's perceptions of discrimination and the U. S. presidency. *Analyses of Social Issues and Public Policy*, *8*, 1–30.

Bigler, R. S., Averhart, C. A., & Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, *39*, 572–580.

Bigler, R. S., Brown, C. S., & Markell, M. (2001). When groups are not created equal: Effects of group status on the formation of intergroup attitudes in children. *Child Development*, *72*, 1151–1162.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, *68*, 530–543.

Bigler, R. S., & Liben, L. S. (1993). A cognitive-developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*, *64*, 1507–1518.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*, 162–166.

Blake, R., & Dennis, W. (1943). The development of stereotypes concerning the Negro. *Journal of Abnormal and Social Psychology*, *38*, 525–531.

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about in-groups and out-groups: Reconceptualizing research on prejudice. *Personality and Social Psychology Review*, *5*, 118–128.

Cameron, L., Rutland, A., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training. *International Journal of Behavioral Development*, *31*, 454–466.

Corenblum, B. (2003). What children remember about ingroup and out-group peers: Effects of stereotypes on children's processing of information about group members. *Journal of Experimental Child Psychology*, *86*, 32–66.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*, 213–244.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. Dovidio & S. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). San Diego, CA: Academic Press.

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition*, *54*, 209–252.

Hitlin, S., Brown, J. S., & Elder, G. H. Jr. (2006). Racial self-categorization in adolescence: Development and social pathways. *Child Development*, *77*, 1298–1308.

Hughes, D. (1997). Racist thinking and thinking about race: What children know but don't say. *Ethos*, *25*, 117–125.

Kao, G. (2000). Group images and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum*, *15*, 407–430.

Katz, P. A., & Zalk, S. R. (1978). Modification of children's racial attitudes. *Developmental Psychology*, *14*, 447–461.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, F31–F36.

Killen, M., & Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer groups contexts. *Child Development*, *72*, 174–186.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*, *70*, 1163–1180.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, *74*, 1421–1436.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied out-group: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31*, 34–37.

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*, 250–259.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., et al. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised. *Personality and Social Psychology Bulletin*, *27*, 996–1010.

Maykovich, M. (1972). Reciprocity in racial stereotypes: White, black, and yellow. *American Journal of Sociology*, *77*, 876–897.

McGuire, W. J., & McGuire, C. V. (1988). Content and process in the experience of self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 97–144). New York: Academic Press.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*, 498–515.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical processing* (pp. 179–195). New York: Cambridge University Press.

Nesdale, D. (1999). Social identity and ethnic prejudice in children. In P. Martin & W. Noble (Eds.), *Psychology and society* (pp. 92–110). Brisbane: Australian Academic Press.

Nesdale, D. (2001). The development of prejudice in children. In M. A. Augoustinos & K. J. Reynolds (Eds.), *Understanding Prejudice, Racism, and Social conflict* (pp. 57–73). London: Sage.

Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 219–245). New York: Psychology Press.

Nesdale, D., & Flesser, D. (2001). Social identity and development of children's group attitudes. *Child Development*, *72*, 506–517.

Niemann, Y. F., Jennings, L., Rozelle, R. M., Baxter, J. C., & Sullivan, R. (1994). Use of free responses and cluster analysis to determine stereotypes of eight groups. *Personality and Social Psychology Bulletin*, *20*, 379–390.

Olson, K. R., & Dweck, C. S. (2008). A blueprint for social cognitive development. *Perspectives on Psychological Science*, *3*, 193–202.

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development*, *77*, 847–860.

Pfeifer, J. H., Ruble, D. N., Fuligni, A. J., Bachman, M. A., Alvarez, J. M., & Cameron, J. A. (2007). Social identity and intergroup attitudes in immigrant and non-immigrant children. *Developmental Psychology*, *43*, 496–507.

Phinney, J. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, *7*, 156–176.

Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information. *Journal of Personality and Social Psychology*, *80*, 876–893.

Rowley, S. J., Kurtz-Costes, B., Mistry, R., & Feagans, L. (2007). Social status as a predictor of race and gender stereotypes in late childhood and early adolescence. *Social Development*, *16*, 150–168.

Ruble, D. N., Alvarez, J. M., Bachman, M. A., Cameron, J. A., Fuligni, A. J., Garcia Coll, C., et al. (2004). The development of a sense of "we": The emergence and implications of children's collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 29–76). East Sussex, UK: Psychology Press.

Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, *78*, 1121–1136.

Rutland, A., Cameron, L., Bennett, L., & Ferrell, J. (2005). Interracial contact and racial constancy: A multi-site study of racial intergroup bias in 3-5 year old Anglo-British children. *Applied Developmental Psychology*, *26*, 699–713.

Semaj, L. (1976). The development of racial-classification abilities. *Journal of Negro Education*, *50*, 41–47.

Semaj, L. (1980). The development of racial evaluation and preference: A cognitive approach. *Journal of Black Psychology*, *6*, 59–79.

Taylor, M. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, *67*, 1555–1571.

Taylor, M., Rhodes, M., & Gelman, S. (2009). Boys will be boys; cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development*, *80*, 461–481.

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 88–114). Hillsdale, NJ: Erlbaum.

Yzerbyt, V. Y., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review*, *5*, 141–155.

Yzerbyt, V. Y., Rocher, S. J., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. Oakes, N. Ellemers, & A. Haslam (Eds.), *The psychology of stereotyping and group life* (pp. 20–50). Oxford, UK: Blackwell.

Zeligs, R. (1947). Children's intergroup concepts and stereotypes. *Journal of Educational Sociology*, *21*, 113–126.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1814–1827

Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping

Margarita Svetlova, Sara R. Nichols, and Celia A. Brownell
University of Pittsburgh

The study explored how the meaning of prosocial behavior changes over toddlerhood. Sixty-five 18- and 30-month-olds could help an adult in 3 contexts: instrumental (action based), empathic (emotion based), and altruistic (costly). Children at both ages helped readily in instrumental tasks. For 18-month-olds, empathic helping was significantly more difficult than instrumental helping and required greater communication from the adult about her needs. Altruistic helping, which involved giving up an object of the child's own, was the most difficult for children at both ages. Findings suggest that over the 2nd year of life, prosocial behavior develops from relying on action understanding and explicit communications to understanding others' emotions from subtle cues. Developmental trajectories of social-cognitive and motivational components of early helping are discussed.

Human children begin to behave prosocially very early in life, before 2 years of age. Studies have documented 1-year-olds' abilities to comfort others in distress, participate in household tasks, and help adults by bringing or pointing to out-of-reach objects (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Rheingold, 1982; Warneken & Tomasello, 2006, 2007; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). These apparently prosocial acts, emerging so early in ontogeny, have been of interest to social, developmental, and evolutionary psychologists as possible precursors to such human-specific characteristics as concern for others, cooperation and, ultimately, altruism (Fehr & Rockenbach, 2004; Piliavin & Charng, 1990; Stevens & Hauser, 2004; Warneken & Tomasello, in press). The origins of these early behaviors, however, are not entirely understood, and the mechanisms underlying them remain a matter of debate. If prosocial responding in adults is driven by understanding others' emotions and desires in combination with a motivation to act on others' behalf (Batson, Ahmad, Powell, & Stocks, 2008), how can we explain prosocial acts in very young children who have still rudimentary social-cognitive abilities and emergent other-oriented motivational systems?

Current theory and research provide different, sometimes contrasting, views about contributors to the early development of prosocial behavior. Some have proposed that infants' interest in people and their actions combined with basic affiliative and imitative tendencies lie at the root of their initial prosocial responses (Grusec, 2006; Hay & Cook, 2007; Rheingold, 1982). Others maintain that the critical contributor to early prosocial behavior is the developing ability to differentiate another's internal states from one's own and to relate one's own emotions and needs to another's in order to act on the other's behalf (Bischof-Kohler, 1991; Mascolo & Fischer, 2007; Moore, 2007; Zahn-Waxler & Radke-Yarrow, 1990). Still others argue for an innate biological substrate for empathy and altruism in infants (Hoffman, 1975; Tomasello, 2008; Zahn-Waxler, Robinson, & Emde, 1992), in part because some prosocial behaviors appear to be uniquely human and emerge before formal socialization begins, suggesting that children may start out being naturally altruistic (Silk et al., 2005; Warneken & Tomasello, 2009).

This research was supported in part by grants from the National Institute of Child Health and Human Development, HD043971 and HD055283, to the last author. Portions of this research were presented at the International Conference on Infant Studies. We thank all the children and parents who volunteered their time to participate in this study. Special thanks to Ranita Anderson, Catie Henderson, Maryam Khatami, Liz Posti, Stephanie Sivers, Hilarie Stern, and Emily Yarrison for assistance with recruitment, data collection, and coding.

Correspondence concerning this article should be addressed to Margarita Svetlova, Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260. Electronic mail may be sent to mas53@pitt.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0015

These different theoretical perspectives place different emphasis on the social-cognitive and motivational components of early prosocial responding. In addition, the *nature* of each of these components may differ as a function of age, giving the same behavior (e.g., bringing something to someone) different meaning depending on the child's social-cognitive and motivational competences. Prosocial behavior of an infant or toddler may be driven by different forms of social understanding than the same behavior exhibited by an older child or an adult. For example, infants may first interpret others' emotional expressions and behaviors as being about the objects to which they are directed rather than about the other person's mental states or psychological attitudes (Gergely, Egyed, & Kiraly, 2007). Correspondingly, young children's social responses may be based on less sophisticated inferences about the minds of others, or on even simpler mechanisms such as reading behavioral cues or understanding the contingencies between behavior and its outcomes (e.g., Moore & Povinelli, 2007; Perner & Ruffman, 2005).

We can ask, then, to what extent do 1-year-olds who help their parents empty the dishwasher, wash the dog, sweep up spilled cereal, and the like, do so out of caring concern for the parent's needs as opposed to being interested in the action itself, enjoying the social exchange, or even relishing the praise that often follows? What inferences do toddlers make about parents' desires, feelings, or needs when they help? When are they able to be altruistic, giving up something of value knowingly and voluntarily? One way to clarify such questions is to compare children's performance in situations that require the same basic behavioral competence but that differ with respect to their underlying understanding and motivation.

During the 2nd year of life, social understanding becomes increasingly "mentalistic" (Flavell, 1999), and many aspects of social behavior are transformed (see Brownell & Kopp, 2007), making this period particularly relevant for investigating such issues. In the current study, we examined helping behavior in 1- and 2-year-olds, a period when children's social understanding rapidly develops—and when prosocial responding first emerges. We presented toddlers with tasks that were similar behaviorally but differed in their specific social-cognitive and motivational demands. Our premise was that if children behave prosocially in versions without the relevant demand and fail to do so in corresponding versions with that demand, we would be able to isolate specific components of prosocial responding that are and are not yet in place at a given age. The factors that we experimentally manipulated were the nature of the inference that the child had to make about the other's need, the cost of helping to the child, and the amount of communicative support available to the child about the other's need and how to help. By contrasting the conditions under which 1-year-olds help with those under which 2-year-olds help, we aimed to shed light on developmental changes in the sources of early prosocial behavior.

Instrumental Versus Empathic Helping

The 2nd year of life is witness to major developmental change in prosocial behavior. Instrumental helping, or assisting another in achieving an action-based goal such as searching for or getting something out of reach, appears by 12–14 months of age (Liszkowski et al., 2006; Warneken & Tomasello, 2007), presumably deriving from early-developing understanding of goals and goal-directed behaviors (Woodward, 1998). The ability to respond prosocially to others' emotional distress emerges somewhat later, with a significant increase between 18 and 24 months of age in expressions of concern and comforting behavior toward others in pain (Zahn-Waxler, Radke-Yarrow, et al., 1992). This is associated with the development of objective self-awareness and likely to be a function of children's corresponding understanding of others as psychological agents (Hoffman, 2007; Moore, 2006, 2007; Zahn-Waxler, Radke-Yarrow, et al., 1992).

In addition to growing social understanding, the motivational mechanisms involved in early prosocial behavior may undergo developmental change during toddlerhood. Not all the possible motivations for behaving prosocially are other oriented, perhaps especially in young children. For example, in a study by Rheingold (1982), 18-month-olds participated at high rates in "helping" their parents to set a table, sweep up bits of paper, put cards in a box, and so forth. However, the fact that children in the study often initiated an action even before the adult expressed a need or started to act points to the possibility that they may have simply enjoyed participating in social activities alongside adults, were seeking adults' approval, or found the activities themselves interesting or fun, rather than being concerned about the other's well-being. Even in older children and adults, prosocial behavior can be motivated by self-oriented concerns such as a desire for social approval, concrete rewards, or reciprocal prosocial responding (Eisenberg, 2005). It is possible, then, that very early instances of *instrumental helping* are motivated as much by more rudimentary forms of self-interest, or interest in the persons, objects, or actions involved, as by compassionate concern for others. *Empathic helping*, that is, prosocial responding grounded in other-oriented concern, may develop later along with children's growing understanding of others as independent agents with internal states that differ from children's own (Hoffman, 2007; Mascolo & Fischer, 2007).

One purpose of the current study was to test this hypothesis by extending previous work on toddlers' instrumental helping (Rheingold, 1982; Warneken & Tomasello, 2006) to include empathic helping and to test 2-year-olds as well as 1-year-olds. We built on the existing experimental work by Warneken and Tomasello (2006), in which they demonstrated that 1-year-old children readily help an adult instrumentally with a variety of goal-directed actions and do so more often than in comparable control tasks in which helping is not necessary. In the current study, we contrasted instrumental and empathic helping by presenting 18- and 30-month-old children with tasks in which they could help an adult by offering her an object that she needed, for example, bringing her an out-of-reach toy so she could continue an interrupted game, or giving her a blanket to make her warm. We expected that children would be more likely to demonstrate prosocial behavior when it involved helping someone to complete an interrupted action (instrumental helping) than when helping was based on inferring and alleviating another's negative internal state (empathic helping). We also expected that this difference would be greater for 18-month-olds than for 30-month-olds, who are more skilled at inferring others' internal states (Lewis & Carpendale, 2002; Repacholi & Gopnik, 1997) as well as more concerned about them (Radke-Yarrow, 1986; Spinrad & Stifter, 2006; Zahn-Waxler, Radke-Yarrow, et al., 1992).

Altruistic Helping

Prosocial behavior, even when it is other oriented, may not always be genuinely altruistic. A child can give a hug to a distressed friend or bring Mom her mittens, which may improve the other person's well-being without diminishing the child's own well-being; in fact, such behaviors are often rewarded, for example, by a reciprocal hug or praise. To what extent are these behaviors altruistic? The answer to this question depends in part on how we define altruism. Scholars from different disciplines provide different definitions: In evolutionary biology, a behavior is deemed altruistic only when it raises the expected reproductive success of the recipient at the expense of the reproductive success of the donor (Kitcher, 1998), whereas psychologists generally use less strict terms, although their definitions, too, range in the degree of self-sacrifice required. Grusec, Davidov, and Lundell (2002) characterize altruistic behavior as intentional assistance to others that comes at a cost to the donor. Piliavin and Charng (1990) argue that altruistic behavior must be performed without expecting any external reward. The general agreement seems to be that for an act to be considered altruistic, the recipient's well-being must be the donor's ultimate goal (Batson et al., 2008; Foster, Wenseleers, & Ratnieks, 2006); that is, the recipient's benefit must be perceived as an end in itself, not as a means to the achievement of some of the donor's other goals (Kitcher, 1998). When does the ability to act on behalf of others—without considering one's own interests and potentially at a cost to oneself—first emerge?

Young children may behave prosocially before they behave altruistically. Toddlers are notoriously poor at sharing their possessions, but with the rapid development of self- and other-understanding over the 2nd year their rates of sharing, helping and cooperating increase (Hay, 2006; Hay & Cook, 2007; Smiley, 2001). However, it is possible that costly helping, especially when it involves sacrificing one's own possessions, may become more difficult, not less so, as children better understand ownership and the costs of sharing in the 3rd year of life (Fasig, 2000; Hay, Caplan, Castle, & Stimson, 1991; Imbens-Bailey & Pan, 1998). Thus, children's motivation to behave altruistically, subordinating their own interests to those of others, may instead decrease with advances in social understanding.

Although toddlers may be motivated to help others, few empirical studies have directly tested their willingness to help as a function of whether self-sacrifice is required or not (cf. Warneken & Tomasello, 2008). Thus, the second purpose of the current study was to examine children's helping behavior in situations that involved someone else's belongings in comparison to situations that required the children to give up their own belongings altruistically (e.g., giving their own blanket brought from home to an adult who was feeling cold). Prior research has shown that children will sometimes give up something in their possession (e.g., Vaish, Carpenter, & Tomasello, 2009), or willingly stop playing an interesting game (Warneken & Tomasello, 2008) to help an adult in distress. Giving up something of one's own, especially an object from home, is still a different type of cost, and arguably a more stringent test of altruistic motivation insofar as it carries greater personal significance and value (Blake & Rand, 2010). We expected children at both 18 and 30 months of age to help more often and more quickly in tasks that did not require them to sacrifice their own belongings. We also expected older children to be more altruistically motivated, hence to help more readily than younger children even in the face of a personal sacrifice; however, we entertained the possibility suggested by Hay et al. (1991) that older children could be less willing to give up something of their own.

The Role of Adult Communication in Early Prosocial Behavior

Prosocial behavior develops within the crucible of social interaction. It is through dyadic interaction with adults that young children's social understanding and responding are socialized and shaped (Dunn, 1988; Grusec, 2006). Notably, in most of the studies reviewed earlier adults used a variety of ostensive cues to communicate their emotions, intentions, or desires explicitly, often with exaggeration to make them salient and clear. Such communication may serve to scaffold children's attention to, interest in, and inferences about others' internal states and the characteristics of the objects or events toward which adults behave. Some researchers have found that infants require the supportive, mutual engagement of an adult to recognize what the adult is attending to (Moll, Carpenter, & Tomasello, 2007) and that infants' ability to interpret others' behavior and to understand what another person feels, needs, or wants may initially depend on adults' scaffolding and communicative support (Carpendale & Lewis, 2004; Ruffman, Slade, Devitt, & Crowe, 2006; Tomasello, 1995). In a recent study of sharing in toddlers, it was shown that 24-month-olds were able to infer and act on an adult's desire for a snack from the adult's explicit verbal expressions of desire, but not when the adult was silent (Brownell, Svetlova, & Nichols, 2009). Prosocial behavior may therefore be more likely when adult communications make the other's needs, desires, and emotions more apparent and reduce the need for complex inferences about others' internal states.

The third purpose of our study was to examine how adults' communications support young children's prosocial responding. To this end, we systematically manipulated the expressive cues provided by the adult during the helping tasks. For each task, the adult made her need and how to help progressively more explicit with a fixed sequence of specific gestural, vocal, and verbal cues, for example, first simply vocalizing "Brrrr!" with shivering and shaking, then reaching toward an out-of-reach blanket with gaze alternation between the child and the blanket. The cues culminated in a direct request for a particular object (e.g., "Can you bring me the blanket?"). Children were assigned a helping score based on when in the sequence of cues they helped the adult. If they helped spontaneously without any supportive cueing from the adult, they received the maximum score, whereas they received the minimum score if they helped only when specifically requested to do so. We expected that 18-month-olds would receive lower scores than 30-month-olds, that is, that younger children's helping responses would depend on more explicit and direct communications about the adult's needs. We also expected that the communicative support from adults would become more necessary as the complexity of the inference about the other's internal states increased; thus, we expected that children would receive higher scores for instrumental helping than for empathic helping and higher scores for empathic helping than for altruistic helping.

Overview of Design

To examine the roles of developing social understanding, prosocial motivation, and adult communicative support in early prosocial behavior, we presented each child with nine tasks in which an adult needed help and the child could provide it by bringing a nearby object that the adult could not reach. We contrasted children's *instrumental* helping behavior in an "action" condition, in which the adult needed an object to complete a goal-directed action, with their *empathic* helping in an "emotion" condition, in which the adult needed an object to alleviate a negative emotion. Even though the target behavior was identical in the two conditions, we expected children to help more readily in the action condition than in the emotion condition. Because understanding of goal-directed behavior developmentally precedes the ability to infer and represent others' internal emotional states, we expected greater difference between the two conditions among 18-month-olds than among 30-month-olds. We also contrasted children's instrumental and empathic helping with *altruistic* helping in which they had to give up something of their own to help the adult (the "altruism" condition). We expected children to help more readily when they did not have to sacrifice their own belongings. Finally, in each helping task we manipulated the adult's communications about her need and what could be done to help her. We expected younger children to require more communicative cues to support their helping behavior, especially in the emotion and the altruism conditions.

Method

Participants

Thirty-two 18-month-olds (*M* = 18.46 months, *SD* = 0.48; 15 boys and 17 girls) and thirty-three 30-month-olds (*M* = 30.32 months, *SD* = 0.68; 18 boys and 15 girls) participated in the study. Seven additional children were tested but their data were not usable because of procedural error or the child's refusal to participate in any of the tasks. Participants came from working- and middle-class families from a medium-sized city and surrounding suburbs. Seventy-eight percent were Caucasian, 11% were Asian, 7% were African American, and 4% were Hispanic. All children were walking, talking, and were healthy and developing normally by parent report.

General Procedure

Children arrived to a laboratory playroom with their caregivers who remained in the room completing questionnaires and were asked not to encourage, reward, or comment on the child's behavior. Testing was conducted by a female experimenter (E) with the help of a female assistant (AE). The session began with a warm-up and familiarization period during which E and AE played with the child for several minutes to ensure that the child was comfortable approaching and interacting with both of them. During warm-up, E and AE also demonstrated various objects that were later used during testing, to be sure that children understood their use and to control for possible differential exposure prior to the study. Specifically, AE came in with messy hair and showed how to fix it with a hairclip, also giving the child a chance to play with the hairclip. She then gave a small stuffed toy with an attached hairclip to the child and told the child that both the toy and the hairclip were for him or her to take home (the toy accompanied the hairclip to make the gift equally interesting for girls and boys). AE also showed the child how a blanket could be wrapped around her shoulders to make her feel warm. E showed her teddy bear to the child, saying that it was her special toy that made her feel happy. Parents had been asked to bring the child's own blanket and favorite toy from home; thus, together with the newly acquired hairclip, there were three objects belonging to the child. Before the experiment started, AE placed all the objects behind a curtain; the objects were later retrieved as relevant in each testing episode. The experimental session was structured so that testing episodes were alternated with free-play episodes. The session was video recorded through a one-way mirror for later coding.

Nine helping tasks (one trial per task) were presented, in which E demonstrated difficulty or distress which could be alleviated by giving her a particular object. Prior to each trial, the target object (the one that E needed) and two distracter objects were placed on a tray out of E's reach, but within the child's reach. The distracter objects served as targets in other trials. On each trial, E provided up to eight progressively more explicit cues as to her need or emotion and what could help or comfort her. All tasks were administered within subjects. To control for possible fatigue or familiarization effects, the order of tasks was counterbalanced using a modified Latin square design, with a total of nine possible orders; no two tasks from the same condition ever occurred in a row. Approximately equal numbers of children received each order.

Conditions and Tasks

The nature of the inference that children had to make about E's need and the degree of sacrifice they had to make were manipulated via three conditions (action, emotion, and altruism), with three trials per condition. In the action condition, the child had to infer E's action-related goal (three trials); in the emotion condition, the child had to infer E's emotion or internal state (three trials). In both of these conditions, the needed object belonged to E. The altruism condition (three trials) was identical to the emotion condition except that the child had to give up something of his or her own to help or comfort E, thus requiring a sacrifice on the child's part to alleviate E's distress.

More specifically, in the action condition, E had difficulty completing three different goal-oriented actions involving objects that were out of reach because they had been dropped or misplaced; the child could help by bringing the object necessary to complete each action. This condition emphasized the interrupted action, not E's internal state. In the emotion and altruism conditions, E demonstrated three different negative internal states (sadness, cold, and frustration); the child again could help by bringing specific objects to E. These conditions emphasized the adult's distress and the child's help served to alleviate the distress instead of completing an interrupted action.

The three trials in each condition were equated across conditions to ensure that the tasks were parallel, that is, involved the same behaviors and objects in each condition. The three task types in each condition were: (a) need for a clip for attaching something, (b) need for a blanket or cloth for wrapping, and (c) need for a toy for playing or comforting. Thus, the tasks were similar across conditions in terms of the target object needed and the specific helping behavior required, but they had different social-cognitive and motivational demands depending on the condition. For example, a clip could be handed to the experimenter to help her complete the action of clipping cloths to a line (action condition) or to alleviate her feeling of frustration with her messy hair (emotion and altruism conditions), with the clip belonging to the child in the altruism condition. Hence, the required helping behavior and the required object were identical in the three conditions (e.g., handing a clip to E), but helping in each condition required a different inference about the need of the recipient and/or a different motivation. Whereas in the action condition helping was based on understanding another person's goal and motivated by wanting to facilitate an interrupted goal-directed action, in the emotion condition it required understanding another person's internal state and wanting to alter that state, with an additional altruistic motivation in the altruism condition.

Description of Tasks

Action Condition

Clipping task (adapted from Warneken & Tomasello, 2006). E is clipping square pieces of fabric to a clothesline and drops a clothespin out of reach. Child's target behavior: Hand the clothespin to E.

Wrapping task. E is wrapping wooden blocks in square pieces of fabric (sitting behind a small table), and runs out of wrappers on the final block. An additional wrapper is available on the tray which can be reached by the child but not by E. Child's target behavior: Hand the wrapper to E.

Toy task. E is playing with a toy car together with a toy driver that fits in the car, then absentmindedly places the car on the tray near the child to play another short game with the child. She then reseats herself behind the table and away from the tray picks up the driver to start playing again, but has no car. Child's target behavior: Hand the car to E.

Emotion Condition

Clipping task. AE places the hairclip on the tray near the child and leaves. E enters the room with her hair in her eyes, sits on the floor out of reach of the tray, and demonstrates her frustration and distress with actions, moans, and sighs as she tries unsuccessfully to move her messy hair away from her face. Child's target behavior: Hand the hairclip to E.

Wrapping task. E shows her blanket to the child, reminding the child that it makes her warm. She then places the blanket on the tray near the child as she begins a game with the child, moves out of reach of the tray, and sits on the floor. E suddenly becomes cold, shivering, rubbing her arms, and saying "Brrrr" with a distressed expression on her face. Child's target behavior: Hand the blanket to E.

Toy task. E shows her teddy bear to the child, reminding the child that it is her favorite toy and that it makes her happy. She then places the teddy bear on the tray near the child and moves out of reach of the tray. AE enters and whispers something to E, who immediately becomes sad, sighing and sobbing as she sits on the floor. Child's target behavior: Hand the teddy to E.

Altruism Condition

Clipping task. AE places the child's hairclip (given to the child during the warm-up and attached to a stuffed toy) on the tray near the child and leaves. E enters the room with her hair in her eyes, sits on the floor out of reach of the tray, and demonstrates frustration with actions, moans, and sighs as she tries unsuccessfully to move her messy hair away from her face. Child's target behavior: Hand the child's *own* hairclip to E.

Wrapping task. E places the child's blanket, brought from home, on the tray near the child; she points it out and talks about it with the child, mentioning that it must be warm. She then begins a game with the child, moving out of reach of the tray, and sits on the floor. E suddenly becomes cold, shivering, rubbing her arms, and saying

"Brrrr" while looking distressed. Child's target behavior: Hand the child's *own* blanket to E.

Toy task. E places the child's favorite toy, brought from home, on the tray near the child; she points it out and talks about it with the child, saying that it looks very special and must make the child happy. E then moves out of reach of the tray and suddenly receives a call on her cell phone, which makes her sad, and she begins to sigh and sob. Child's target behavior: Hand the child's *own* toy to E.

Communicative Cues

On each trial, the experimenter provided up to eight progressively more explicit cues about her need or emotion and what the child could do to help or comfort her. The cues and their order of presentation were the same across trials. The first two cues communicated the experimenter's *state* gesturally, then verbally (e.g., shivering; saying "I'm cold"); the third cue stated the *general need* (e.g., "I need something to make me warm"); the fourth cue drew the child's attention to the *target object* by labeling it (e.g., "a blanket!"); and the rest of the cues informed the child with increasing specificity about *how to help or comfort* the experimenter, with the final cue giving a very *specific instruction* (e.g., "Can you bring me the blanket?"). Table 1 displays the order of presentation and description of each of the eight cues.

Each cue was presented for 5–7 s. Once the child handed the target object to the experimenter, she stopped providing cues. If the child brought one of the distracter objects first, E took it from the child with a neutral facial expression and continued to provide cues until the child brought the target object. If the child did not bring the target object by the last cue, the experimenter got up and retrieved it herself without reaction or comment and then proceeded to the next play episode. To reduce the possibility of simple compliance or attempts to seek adult approval, the experimenter did not thank, praise, or reward the child when the child brought the object, but instead neutrally described the result (e.g., "Now I can clip," "Now I feel warm").

Measures

Children's responses to each helping task were coded from video records. For each task, the child received a "target helping score," corresponding to the specific cue in the sequence of eight communicative cues when the child brought the target object to the experimenter (see Table 1 for the scores assigned to each cue). Higher scores represent more skilled responding, that is, earlier responses in the cue sequence, requiring less social and communicative support. A child who did not bring the target object by the final cue on a given task received a score of 0 for that task. As an indicator of their general propensity to help, children were also coded for whether they brought any object, either target or distracter, at any cue level, for each task (coded 0 or 1). Video records were coded by two trained research assistants blind to the study's hypotheses. Reliability was calculated on 21% of the data, resulting in a weighted $k = .89$.

Results

Preliminary analyses were conducted to check for gender, task, and task order effects. No differences

Table 1
Communicative Cues and Corresponding Helping Scores

Order of presentation	Description	Helping score assigned
1	Facial/bodily/vocal expression of general need (e.g., hands up, looking around, "hmmm") or internal state (e.g., shivering with cold, rubbing and hugging oneself to get warm, "brrrr")	8
2	Naming the interrupted action (e.g., "I can't clip!") or internal state (e.g., "I am sad")	7
3	Verbal expression of a general need for an object (e.g., "I need something to clip with" or "I need something to make me feel warm")	6
4	Naming the specific object that would meet the need (e.g., "A clothespin!" or "A blanket!")	5
5	Alternating gaze between the object and the child, as a nonverbal request to get the object	4
6	Reach and gesture toward the object, as a more explicit request to get the object	3
7	General verbal request for help ("Can you help me?")	2
8	Specific verbal request (e.g., "Can you bring me the blanket?")	1

were found for gender or task order on any outcome variable, so all subsequent analyses were conducted on data collapsed over gender and order. Overall, task effects varied unsystematically across conditions, suggesting that no particular type of task was more or less familiar or difficult than others; in particular, tasks did not differ in difficulty within the emotion condition. Analyses also showed that task did not interact with order or gender to affect performance, nor did order interact with condition.

Two sets of substantive analyses are presented. The first set addressed general questions about whether 30-month-olds helped more often across contexts and manipulations than did 18-month-olds, and whether their helping was more specific and appropriate. The second set of analyses examined age and condition differences in how readily children helped and how much communicative support they needed, using the target helping score as the dependent measure.

General Propensity to Help and Appropriateness of Help

To examine age differences in children's general propensity to help the experimenter, we analyzed the number of times children brought any object (either the target or the distracter object) across all nine trials, regardless of the cue level at which they did so. All children exhibited some degree of helping toward the experimenter. Among 18-month-olds, every child helped by bringing some object on at least two of the nine trials, and 44% helped on all nine. Among 30-month-olds, every child helped on at least six of the nine trials, and 84% helped on all nine. An independent-samples t test on the mean number of helping responses yielded a significant age difference, with 30-month-olds helping more often ($M = 8.74$, $SD = 0.68$) than 18-month-olds ($M = 7.40$, $SD = 1.88$), $t(61) = 3.72$, $p < .001$. Thus, older children were more inclined to help overall, but even younger children helped, on average, on more than seven of the nine opportunities. It is important to note, however, that this includes instances of directed or instructed helping when the experimenter specifically requested the target object by name (e.g., "Can you bring me the blanket?"), as well as instances of helping by bringing the distracter object, that is, without discriminating whether the object was the one that the experimenter actually needed.

To examine children's ability to help appropriately by bringing the specific object that the adult needed, we conducted a two-way repeated measures analysis of variance (ANOVA) with age group (18 and 30 months) as the between-subjects factor, and condition (action, emotion, altruism) as the within-subjects factor, on the mean number of times children brought the target object, regardless of cue level (see Figure 1). Across conditions, older children brought the target object significantly more often ($M = 2.87$ out of three trials per condition, $SD = 0.38$) than did younger children ($M = 2.31$, $SD = 0.51$), $F(1, 61) = 14.73$, $p < .001$. There was also a significant effect of condition, $F(2, 122) = 19.16$, $p < .001$. Post hoc pairwise comparisons (LSD) revealed that children of both ages brought the target significantly more often in the action condition ($M = 2.84$, $SD = 0.41$) than in the emotion condition ($M = 2.69$, $SD = 0.67$), $p = .03$, and significantly more often in the emotion condition than in the altruism condition ($M = 2.24$, $SD = 0.98$), $p = .001$. The interaction was not significant. Thus, collapsing across cue levels, 30-month-olds more often helped and more often helped appropriately by bringing the object that the experimenter actually needed. All children tended to exhibit more appropriate helping in the action condition than in the emotion and the altruism conditions.

How Readily Children Helped as a Function of Age and Helping Condition

To determine whether there were age or condition differences in how readily children helped, that is, how much communicative support they needed to help appropriately, a two-way repeated measures ANOVA, with age group (18 and 30 months) as the between-subjects factor and condition (action, emotion, altruism) as the within-subjects factor, was conducted on the target helping scores. Higher scores corresponded to earlier, more

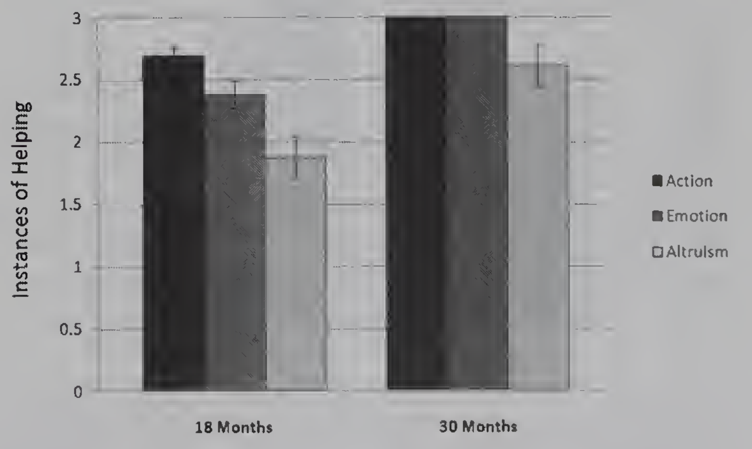


Figure 1. Mean number of times (± *SE*) children helped by age and condition (out of three possible instances per condition).

skilled helping with less communicative support from the adult. Nonparametric analyses for ordinal data produced identical results.

A significant difference between age groups was found, $F(1, 61) = 61.99$, $p < .001$. Across conditions, 30-month-olds helped more readily, that is, at earlier cue levels ($M = 5.41$, $SD = 1.14$), than did 18-month-olds ($M = 3.25$, $SD = 1.07$). There was also a significant effect of condition, $F(2, 122) = 68.04$, $p < .001$. Post hoc comparisons (LSD) revealed that across ages, children helped more readily in the action condition ($M = 5.62$, $SD = 1.70$) than in the emotion condition ($M = 4.28$, $SD = 2.01$), $p < .001$, and in the emotion condition than the altruism condition ($M = 3.09$, $SD = 1.77$), $p < .01$.

These main effects were qualified by a significant interaction between age and condition, $F(2, 122) = 5.91$, $p < .001$ (see Figure 2). Younger children needed significantly less communicative support to help in the action condition ($M = 4.64$, $SD = 1.75$) than in either the emotion condition ($M = 2.79$, $SD = 1.55$), $t(31) = 6.26$, $p < .001$, or the altruism condition ($M = 2.33$, $SD = 1.44$), $t(31) = 7.08$, $p < .001$, which did not differ from each other, $t(31) = 1.75$, *ns*. Older children, on the other hand, helped more readily in the action condition ($M = 6.61$, $SD = 0.88$) than in the emotion condition ($M = 5.77$, $SD = 1.26$), $t(30) = 3.21$, $p = .003$, and more readily in the emotion condition than in the altruism condition ($M = 3.85$, $SD = 1.76$), $t(30) = 5.49$, $p < .001$.

As a more stringent test of children's ability to help, we conducted nonparametric tests on the number of children who helped immediately by bringing the target object after the initial communicative cue from the experimenter (i.e., nonverbal expression of need). Among 18-month-old children, 44% helped the adult immediately and appropriately on one or more of the tasks in the action condition (out of three); that is, they brought the target object when the adult first registered her difficulty with facial expression, vocalization, and/or posture. In the emotion condition, this dropped to 13%, and in the altruism condition only 6% of 18-month-olds helped the adult immediately at least once. Among 30-month-olds, 87% helped immediately one or more times in the action condition, as did 64% in the emotion condition, but only 18% helped immediately at least once in the altruism condition. Age differences in the number of children who helped immediately one or more times were significant for the action condition, $\chi^2 = 20.30$, $p < .001$, and the emotion condition, $\chi^2 = 18.30$, $p < .001$, but not for the altruism condition, $\chi^2 = 2.43$, *ns*.

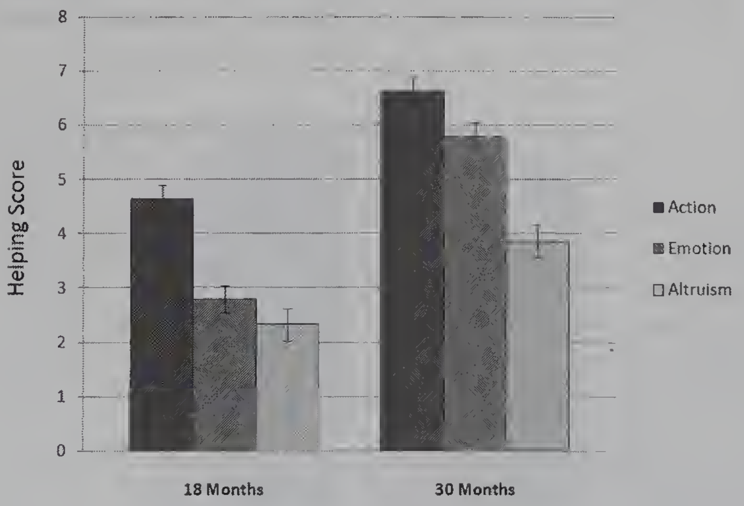


Figure 2. Mean target helping scores (± *SE*) by age and condition.
Note. Higher scores indicate earlier helping with less communicative support.

Discussion

In this study, we examined early developments in young children's ability to help others in need. Our data confirmed previously reported findings that toddlers as young as 18 months are able to help adults instrumentally, and that the propensity to respond prosocially grows significantly between 18 and 30 months of age. The particular novelty of this study was in the manipulation of factors that could underlie helping behaviors at each of the examined ages, specifically, the nature of the inference that the child had to make about the other's need, the cost of helping to the child, and the amount of communicative support provided by the adult in need of help. By varying these factors we aimed to clarify possible social-cognitive and motivational mechanisms that support early prosocial behavior and its development in the 2nd and 3rd years of life. Helping an adult instrumentally to complete an interrupted action-related goal was significantly easier than empathic helping to alleviate an adult's emotional distress, particularly for 18-month-olds. Helping was also easier when it did not require children to give up something that belonged to them, at both 18 and 30 months. We also found that 18-month-olds' helping required significantly greater communicative support and scaffolding from an adult than that of 30-month-olds, whose prosocial behavior was more autonomous and demonstrated greater social understanding.

Age-Related Differences in Helping and Requisite Adult Communication

Prosocial responses emerge early in the 2nd year of life (Rheingold, 1982; Warneken & Tomasello, 2006, 2007) and increase over childhood in concert with growing self-other differentiation and social understanding (Bischof-Kohler, 1991; Brownell & Carriger, 1990; Moore, 2006; Zahn-Waxler, Radke-Yarrow, et al., 1992). However, prosocial responses among 1-year-olds are infrequent, occur more reliably for familiar situations and emotion contexts (e.g., pain; Zahn-Waxler, Radke-Yarrow, et al., 1992) than for less familiar situations (e.g., sadness, fatigue; Sigman, Kasari, Kwon, & Yirmiya, 1992), and can be increased by explicit requests for assistance (Rheingold, 1982; Spinrad & Stifter, 2006). Even at age 2, prosocial behavior remains irregular and does not occur reliably across trials or tasks (Demetriou & Hay, 2004; Hay, 2006).

In the current study, both 18- and 30-month-old children helped the experimenter on the majority of trials. However, these apparently high helping rates included instances in which the adult explicitly requested help from the child and specified what was needed. Furthermore, when we consider the particular circumstances under which 18-month-olds helped, it becomes clear that their performance was mostly accounted for by helping in the action-related tasks, in which the experimenter had difficulty completing a goal-directed action. For emotion-related tasks, where children had to infer the adult's internal state and intervene to alleviate it, 18-month-olds were unlikely to help unless they received substantial communicative support and scaffolding from the adult. On average, the younger group tended to respond after the adult had expressed her internal state or need in different ways and had named the target object. Thirty-month-olds exceeded 18-month-olds on all measures of prosocial behavior in all conditions. As predicted, not only did older children help more frequently, but they also required less communicative support from the adult, tending to help before the adult had named the object that she needed. Thus, 18-month-olds are most able to respond prosocially in situationally obvious, goal-oriented helping situations when they are supported by very clear communications about what the other person needs and what can be done to mitigate the need. By 30 months of age, children's greater ability to infer another's needs from more general information and more subtle cues permits them to recognize another's need for help more easily and to rely less on the adult for this information.

One implication of these findings is that very early prosocial responding depends on social and communicative support from others about their needs and how the child can help. Several investigators of early social-cognitive development, starting with Vygotsky (1978), have emphasized the essential role of partner support when children first acquire new knowledge and skills. In a classic study of 6- to 18-month-old infants and their mothers, Bakeman and Adamson (1984) showed that new forms of social engagement first appeared embedded in and supported by the social interaction context. In their study, mothers initially acted in ways that scaffolded their infants' performance before infants were capable of such performance on their own; with age and repeated exposure, infants themselves exhibited more skilled and autonomous forms of engagement. We can speculate that new forms of social understanding and the corresponding prosocial responses are socialized in a similar way. In the current study, tasks with more challenging social-cognitive demands required more prolonged, specific, and explicit communication; with age, children's need for such communication decreased. While human infants may have a natural predisposition to be prosocial early in development (Warneken & Tomasello, 2009), social and communicative support from adults appears to be important in scaffolding young children's understanding of when and how to enact prosocial responses.

Social Understanding: Actions Versus Emotions

In line with our expectations, instrumental, action-related tasks elicited the most robust helping responses and required the least scaffolding from an adult. Even for 30-month-olds helping in the action condition was significantly easier than in the other two conditions. This is consistent with existing research on development of early social understanding that has shown that by the middle of the 2nd year of life children are able to understand other people's goals and intentions (Meltzoff, 1995; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Even though emotion understanding also undergoes important developments in this period (Eisenberg, 2005; Mascolo & Fischer, 2007), enabling some forms of empathy (Spinrad & Stifter, 2006; Zahn-Waxler, Radke-Yarrow, et al., 1992), empathic helping appears to be more complex than helping with action-related tasks. Unlike helping based on predicting action–effect outcomes in observable actions (Csibra & Gergely, 2007), helping in response to others' emotions requires an additional step in social-cognitive processing, as inferring the goal in this case follows from inferring someone's internal state. Furthermore, emotion-related tasks can lead to personal distress in young children via emotion contagion, perhaps especially when others are in pain or evident distress. This requires young children both to differentiate their own distress from the other's distress and to regulate their own distress if they are to exhibit empathic helping based on other-oriented concern or sympathy rather than based on the wish to reduce their own personal distress (Eisenberg, 2005; Hoffman, 1975). As a result, both cognitive and affective demands are greater in empathic helping compared to instrumental, action-related helping. The ability to understand and assist others with goal-directed actions may represent the first step toward this more complex social understanding and associated prosocial responses.

Together with other recent research, the current study points to an important transition from instrumental helping to empathic helping late in the 2nd year of life and early in the 3rd. Warneken and Tomasello (2007) demonstrated that 14-month-olds are able to help adults in only the most straightforward instrumental tasks, those involving reaching. Between 18 and 24 months of age, toddlers are able to help in a wider array of instrumental tasks that include difficulties with tools, obstacles, means–end relationships and the like, suggesting more advanced understanding of others' action-related goals (Warneken & Tomasello, 2006). In the current study, by contrasting instrumental and empathic helping, we discovered that empathic helping is quite limited at 18 months but relatively skilled by 30 months of age. The developmental picture suggested by integrating these findings is that by 14 months of age, toddlers are able to help others with simple, goal-directed actions; by 18–24 months they can help in action-related situations that are more demanding cognitively but still relatively transparent with respect to another's goals and needs; and by 30 months of age, they are able to help in emotion-related situations requiring more complex inferences about others' needs based on inferring others' feelings and internal states.

Altruistic Motivation: Cost Versus No Cost

The altruism condition, in which helping required the child to give his or her own possession to the experimenter, was included in the design to examine the motivational component of early prosocial behavior. Recall that the only difference between the altruism and emotion conditions was whether the objects needed by the adult were the child's own or the adult's. As the two conditions were otherwise identical, any differences in helping must have been due to affective or motivational, rather than cognitive, mechanisms. As expected, the altruism condition was the most difficult for both ages. Interestingly, however, 18-month-olds' behavior did not differ significantly for the emotion condition versus the altruism condition. Indeed, their helping scores were near the floor in both. In other words, whether a sacrifice was involved or not, 18-month-olds had an equivalently difficult time understanding the adult's emotion-related needs and required substantial input from the adult to come up with a helping response. Among 30-month-olds, however, both the rates of helping and how early in the cue sequence children responded differed significantly between the altruism and emotion conditions, even though the social-cognitive demands were identical. Although 30-month-olds knew about the adult's need and how to help in the emotion tasks, they were less willing to provide costly help in the altruism tasks. For the younger group, therefore, the challenge seemed to be in disentangling the complex social understanding demands of the emotion-related tasks, whereas the older group had the most difficulty with overcoming their possessiveness and giving up something of value.

These findings suggest that very young children exhibit a general prosocial motivation, readily providing instrumental help to an adult when the amount of information about what is needed corresponds to their level of social understanding. Yet, when the demands of the task involve a sacrifice—giving their own possession to alleviate another's distress—this motivation is reduced. Grusec et al. (2002), among others, define prosocial behavior as any intentional action that produces a positive outcome for the recipient, whereas altruism implies that assistance to others comes at some cost to the donor. The fact that instances of costly helping in the current study were quite low and often occurred in response to an adult's direct and explicit request suggests that toddlers' helping responses were unlikely to be genuinely altruistic. Altruistic motivation appears to be a later developing phenomenon, which may build upon the more basic prosocial motivation emerging and developing in toddlerhood.

Several limitations of this study are worth noting. First, the contrast between the action condition and emotion condition was not as strong as it could be in some respects, but in other respects may have been too distinct. For example, the emotion of frustration is in part a response to a blocked goal (Witherington, Campos, & Hertenstein, 2001) and as operationalized in the current study also has an action component (e.g., moving one's hair away from one's face). Clearer distinctions between these conditions might produce clearer developmental differences between instrumental and empathic helping. Another limitation is the absence of an action version of the altruism condition, thereby constraining our examination of altruistic responding to empathy-related helping situations. Future research could address whether altruism emerges earlier in instrumental helping than in empathy-related helping contexts. Finally, perhaps some children failed to help because they chose not to help, believing that adults bear responsibility for helping others (Caplan & Hay, 1989). In particular, it is possible that rates of helping among the older children in the altruism condition were underestimated in the current study if such factors were in operation. One way future research might address this question is by manipulating the circumstances under which helping occurs, such as presence of parents or other adults in the room, or manipulating the value of the to-be-shared object in addition to whether it belongs to the child or not.

In sum, the findings of the current study reveal an important developmental transition at the end of the 2nd year of life when toddlers' helping behavior expands to include empathic as well as instrumental helping. The results point as well to the late emergence of altruistic helping, after other-oriented helping first becomes evident, inasmuch as even 2-year-olds find costly helping especially difficult. This suggests that changes in social understanding and prosocial motivation may be closely linked, with other-oriented concern developing in concert with growth in children's ability to represent and understand others' subjective internal states, and altruistic helping developing later, in concert with understanding of social and moral norms. It would be productive in future research to investigate these links more directly, possibly by including additional measures of self- and other-understanding and empathy, as well as by testing older children in situations that require various types of helping.

References

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55,* 1278–1289.

Batson, C. D., Ahmad, N., Powell, A. A., & Stocks, E. L. (2008). Prosocial motivation. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (pp. 135–149). New York: Guilford.

Bischof-Kohler, D. (1991). The development of empathy in infants. In M. E. Lamb & H. Keller (Eds.), *Infant development: Perspectives from German-speaking countries* (pp. 245–273). Hillsdale, NJ: Erlbaum

Blake, P., & Rand, D. (2010). Currency value moderates equity preference among young children. *Evolution and Human Behavior, 31,* 210–218.

Brownell, C. A., & Carriger, M. S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Development, 61,* 1164–1174.

Brownell, C. A., & Kopp, C. B. (2007). *Socioemotional development in the toddler years: Transitions and transformations.* New York: Guilford.

Brownell, C. A., Svetlova, M., & Nichols, S. (2009). To share or not to share: When do toddlers respond to another's needs? *Infancy, 14,* 117–130.

Caplan, M. Z., & Hay, D. F. (1989). Preschoolers' responses to peers' distress and beliefs about bystander intervention. *Journal of Child Psychology and Psychiatry, 30,* 231–242.

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral & Brain Sciences, 27,* 79–96.

Csibra, G., & Gergely, G. (2007). ''Obsessed with goals'': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica, 124,* 60–78.

Demetriou, H., & Hay, D. F. (2004). Toddlers' reactions to the distress of familiar peers: The importance of context. *Infancy, 6,* 299–318.

Dunn, J. (1988). *The beginnings of social understanding.* Cambridge, MA: Harvard University Press.

Eisenberg, N. (2005). The development of empathy-related responding. In G. Carlo & C. P. Edwards (Eds.), *Moral motivation through the life span* (pp. 73–117). Lincoln: University of Nebraska Press.

Fasig, L. G. (2000). Toddlers' understanding of ownership: Implications for self-concept development. *Social Development, 9,* 370–382.

Fehr, E., & Rockenbach, B. (2004). Human altruism: Economic, neural, and evolutionary perspectives. *Current Opinion in Neurobiology, 14,* 784–790.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21–45.

Foster, K. R., Wenseleers, T., & Ratnieks, F. (2006). Kin selection is the key to altruism. *Trends in Ecology and Evolution, 21*(2), 57–60.

Gergely, G., Egyed, K., & Kiraly, I. (2007). On pedagogy. *Developmental Science*, *10*, 139–146.

Grusec, J. E. (2006). The development of moral behavior and conscience from a socialization perspective. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 243–265). Mahwah, NJ: Erlbaum.

Grusec, J. E., Davidov, M., & Lundell, L. (2002). Prosocial and helping behavior. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 457–474). Malden, MA: Blackwell.

Hay, D. F. (2006). Yours and mine: Toddlers' talk about possessions with familiar peers. *British Journal of Developmental Psychology*, *24*, 39–52.

Hay, D. F., Caplan, M., Castle, J., & Stimson, C. A. (1991). Does sharing become increasingly ''rational'' in the second year of life? *Developmental Psychology*, *27*, 987–993.

Hay, D. F., & Cook, K. V. (2007). The transformation of prosocial behavior from infancy to childhood. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 100–131). New York: Guilford.

Hoffman, M. L. (1975). Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology*, *11*, 607–622.

Hoffman, M. L. (2007). The origins of empathic morality in toddlerhood. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 132–145). New York: Guilford.

Imbens-Bailey, A., & Pan, B. A. (1998). The pragmatics of self- and other-reference in young children. *Social Development*, *7*(2), 219–233.

Kitcher, P. (1998). Psychological altruism, evolutionary origins, and moral rules. *Philosophical Studies*, *89*, 283–316.

Lewis, C., & Carpendale, J. (2002). Social cognition. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 376–393). Malden, MA: Blackwell.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12-and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*, 173–187.

Mascolo, M. E., & Fischer, K. W. (2007). The codevelopment of self and sociomoral emotions during the toddler years. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 66–99). New York: Guilford.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 1–16.

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*, 826–835.

Moore, C. (2006). *The development of commonsense psychology*. Mahwah, NJ: Erlbaum.

Moore, C. (2007). Understanding self and others in the second year. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 43–65). New York: Guilford.

Moore, C., & Povinelli, D. J. (2007). Differences in how 12- and 24-month-olds interpret the gaze of adults. *Infancy*, *11*, 215–231.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*(5719), 214–216.

Piliavin, J. A., & Charng, H.-W. (1990). Altruism: A review of recent theory and research. *Annual Review of Sociology*, *16*, 27–65.

Radke-Yarrow, M. (1986). Affective development in young children. In T. B. Brazelton & M. W. Yogman (Eds.), *Affective development in infancy* (pp. 145–152). Westport, CT: Ablex.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*, 12–21.

Rheingold, H. L. (1982). Little children's participation in the work of adults, a nascent prosocial behavior. *Child Development*, *53*, 114–125.

Ruffman, T., Slade, L., Devitt, K., & Crowe, E. (2006). What mothers say and what they do: The relation between parenting, theory of mind, language and conflict/cooperation. *British Journal of Developmental Psychology*, *24*, 105–124.

Sigman, M. D., Kasari, C., Kwon, J.-H., & Yirmiya, N. (1992). Responses to the negative emotions of others by autistic, mentally retarded, and normal children. *Child Development*, *63*, 796–807.

Silk, J., Brosnan, S., Vonk, J., Henrich, J., Povinelli, D., Richardson, A. S., et al. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature*, *437*, 1357–1359.

Smiley, P. A. (2001). Intention understanding and partersensitive behaviors in young children's peer interactions. *Social Development*, *10*, 330–354.

Spinrad, T. L., & Stifter, C. A. (2006). Toddlers' empathy-related responding to distress: Predictions from negative emotionality and maternal behavior in infancy. *Infancy*, *10*, 97–121.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*, 60–65.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, *45*, 534–543.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Oxford, England: Harvard University Press.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*, 271–294.

Warneken, F., & Tomasello, M. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology*, *44*, 1785–1788.

Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, *100*, 445–471.

Witherington, D. C., Campos, J. J., & Hertenstein, M. J. (2001). Principles of emotion and its development in infancy. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 427–464). Malden, MA: Blackwell.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.

Zahn-Waxler, C., & Radke-Yarrow, M. (1990). The origins of empathic concern. *Motivation and Emotion*, *14*, 107–130.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*, 126–136.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology*, *28*, 1038–1047.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1828–1845

Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups

Daphna Bassok
University of Virginia

Recent studies suggest that the effects of attending preschool vary by race. These findings are difficult to interpret because the likelihood of enrolling a child in preschool also differs across groups. This study used newly released, nationally representative data to examine whether the impact of preschool participation at age 4 varies across racial groups after accounting for selection differences (N = 7,400). Among a subsample of children living below a poverty threshold, no racial differences in preschool impact are detected. However, findings suggest that nonpoor Black children benefit substantially more from preschool than their nonpoor White or Hispanic peers. Implications of these findings are discussed toward understanding the potential of large-scale preschool interventions for narrowing racial achievement gaps.

The achievement gap between White children and Black and Hispanic children is a persistent and widely documented policy concern (Fryer & Levitt, 2004; Reardon & Robinson, 2008). Racial gaps in cognitive ability emerge years before children enter elementary school (Fryer & Levitt, 2006; Haskins & Rouse, 2005; Loeb & Bassok, 2008). A growing body of research provides strong evidence that preschool programs can have positive and long-lasting effects on a host of academic and social outcomes (Belfield, Nores, Barnett, & Schweinhart, 2006; Garces, Thomas, & Currie, 2000; Ludwig & Miller, 2007; Reynolds, Temple, Robertson, & Mann, 2001). Together, these two facts have led to growing consensus that preschool programs are a critical component of any policy agenda aimed at ensuring children—irrespective of their race or socioeconomic status (SES)—can succeed in school.

If preschool interventions are to meaningfully narrow achievement gaps, however, at least one of two conditions must hold. Either programs must be targeted specifically toward low-income or minority populations or the effects of program participation must be larger among these groups. As a growing number of states and communities move away from economically targeted preschool programs, expanding or even universalizing their preschool initiatives, it is particularly important to understand whether program impact varies systematically across groups of children (Barnett, Epstein, Friedman, Boyd, & Hustedt, 2008).

Many studies demonstrate that the effects of preschool are greatest for poor children (Garces et al., 2000; Gormley & Gayer, 2005; Magnuson, Ruhm, & Waldfogel, 2007). Several recent studies suggest that even after controlling for SES, the effects of attending preschool in the year before kindergarten entry are also larger among Black and Hispanic children relative to White children (Gormley & Phillips, 2005; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007). These findings suggest that preschool—even when widely available—might offer a promising strategy for reducing or eliminating racial achievement gaps.

To date, however, research has not examined *why* preschool effects might vary by race. The current study aims to fill this gap, focusing in particular on differential selection processes across groups as one possible explanation. In 2005, 43% of Hispanic children aged 3–5 attended preschool. White and Black children participated at far higher rates, 59% and 66%, respectively (Rooney et al., 2006).

This research was supported by a grant from the American Educational Research Association that receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics of the Institute of Education Sciences (U.S. Department of Education) under NSF Grant DRL-0634035. Additional support was provided by the Stanford Graduate Fellowship. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies. I thank Susanna Loeb, Sean Reardon, Margaret Bridges, Bruce Fuller, Eric Hanushek, Joseph P. Robinson, and Deborah Stipek for invaluable suggestions.

Correspondence concerning this article should be addressed to Daphna Bassok, Curry School of Education, University of Virginia, 405 Emmet Street South, Charlottesville, VA 22904. Electronic mail may be sent to dbassok@virginia.edu.

© 2010 The Author
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0016

These large differences in participation might partially explain the observed differences in effects.

The current study utilizes newly available data from the Early Childhood Longitudinal Study–Birth Cohort (ECLS–B) to examine whether the effects of preschool attendance at age 4 on cognitive outcomes differ by race. Throughout this study preschool refers to any classroom-based program targeted at 4-year-olds including nursery schools, preschool centers, and prekindergarten programs. The ECLS–B is the first study to track a large, nationally representative sample from 9 months to school entry. The availability of data on both family environment and children's cognitive ability at 9 months, which is before most children have started any formal nonrelative care, allows for a more thorough understanding of preschool impacts than has been previously possible using retrospective data.

Differential Preschool Outcomes

Children who attend preschool in the year before kindergarten entry tend to start school with higher cognitive skills than similar children who stay with their parents (Barnett, 1995; Currie, 2001). Experimental studies demonstrate that for very poor Black children attending a high-quality preschool can yield positive effects across a variety of domains well into adulthood (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Schweinhart et al., 2005). Because of their highly targeted populations, however, these studies tell us little about the potential impact of preschool for nonpoor or non-Black children. In order to make comparisons across socioeconomic and racial groups, researchers have turned to nonexperimental data. Large-scale studies show that the effects of attending preschool at age 4 are largest among poor children (Garces et al., 2000; Gormley & Gayer, 2005; Magnuson et al., 2007).

Recent studies suggest that the relation between preschool and child outcomes also differs across racial groups, even after accounting for a host of socioeconomic measures. Using the kindergarten cohort of the ECLS, Loeb et al. (2007) show that the association between preschool participation and reading scores at kindergarten entry is almost 3 times larger among English-proficient Hispanic children than among White children. The effects for Black children are between those of White and Hispanic children.

Similarly, in their quasi-experimental study of universal prekindergarten in Oklahoma, Gormley and Phillips (2005) find that among Hispanic children, preschool participation leads to a 54% improvement on a cognitive assessment compared with no significant effects for White children and a 17% improvement among Black children. Finally, in their study of Head Start, the federally funded preschool program targeted toward low-income children, Currie and Thomas (1999) use within-family comparisons and show that Hispanic Head Start attendees experience particularly large boosts in test scores. They estimate that Head Start participation closes approximately one fourth of the Hispanic–White test score gap.

Explaining Differential "Effects"

The results from these studies suggest that preschool participation might be a powerful tool for reducing racial achievement gaps. However, the mechanism that explains the differences has not been investigated. There are at least five plausible explanations for differences in preschool effects across groups.

First, differences in preschool effects could be driven by differential selection patterns across racial groups. Recall that Hispanic children attend preschool programs at significantly lower rates than White children and that Black children attend at the highest rates. If the Hispanic children who attend preschool are a more "selective" group than their White or Black counterparts then an estimation strategy that fails to explicitly control for these differences would artificially inflate the effects of preschools for Hispanic children. Here "selective" might mean they would have performed better relative to their peers irrespective of care choice but could also mean that relative to other racial groups, Hispanic preschool attendees are more likely to be those who stand to gain the most from preschool.

Previous studies that have explored differences in preschool effects across groups have attempted to adjust for this type of nonrandom selection into preschool. However questions about selection and generalizability remain. For instance, although Loeb et al. (2007) are able to control for the many child and family characteristics available in the ECLS–K, in that study the data were collected *after* the child finished preschool as they were entering kindergarten. There are no measures of pretreatment cognitive outcomes, and no way to definitively check whether observed differences in preschool effects across racial groups are biased by racial differences in pretreatment characteristics. The quasi-experimental research design employed in the Oklahoma study relies on comparisons between children who are quite similar in age but due to their birthdays and to the state's enrollment cutoff for prekindergarten, enroll in the state program 1 year apart. Although the approach provides unbiased estimates of the effects of preschool for children who eventually enroll, it does not address differences between participants and nonparticipants. Because participation rates differ across racial groups, the authors caution readers against making across-group comparisons about program effects.

A second possibility for the observed racial differences in preschool effects is that race is serving as a proxy for other child characteristics, such as poverty or English language ability. The overrepresentation of Black and Hispanic children among groups who may benefit disproportionately from preschool might explain the observed differences in the impact of preschools. As discussed earlier, studies have repeatedly found that the effects of preschool are largest among poor children. Evidence of a differential preschool effect for English language learners is more mixed. Consistent with the hypothesis that preschools are particularly beneficial for children who are learning English, a study of Oklahoma's prekindergarten program shows that Hispanic children whose parents speak Spanish at home benefit far more from preschool than those whose parents primarily speak English (Gormley, 2008). However, other studies have not found stronger preschool effects for nonnative English speakers (Crosnoe, 2007; Currie & Thomas, 1999; Magnuson, Lahaie, & Waldfogel, 2006).

A third possibility is that group differences in preschool experiences might be part of the explanation for differences in effects. If the nature of the preschool intervention differs across racial groups, analyses that fail to account for these key details may lead to ''apples to oranges'' comparisons. For instance, exposure to preschool, in terms of both hours per week and age at entry, varies by race with Black children starting preschool younger and attending for more hours than their White or Hispanic peers (Gormley & Phillips, 2005; Rumberger & Tran, 2006).

Preschool quality also differs across groups. The Cost, Quality and Outcomes Study shows that White children attend preschools that are almost one third of a standard deviation higher on an aggregate quality rating relative to those attended by Black children (Burchinal & Cryer, 2003). Similarly, although 41% of White preschoolers in California have a teacher with an associate's degree or higher in early childhood education, the figure is 34% for Latinos and 13% for Blacks (Karoly, Ghosh-Dastidar, Zellman, Perlman, & Fernyhough, 2008). It is important to note, however, that if these quality differences are driving differences in preschool effects across groups, we would expect Black and Hispanic children to experience *smaller* benefits from care relative to White children. In fact, we observe the opposite.

A final explanation is that the experience of *not* attending preschools may differ across groups. Many studies of preschool compare children who participate to those who use an alternate arrangement—oftentimes parental or relative care (Loeb, Fuller, Kagan, & Carrol, 2004; Magnuson et al., 2007; NICHD Early Child Care Research Network, 2006). The experience of being in this control group may vary across racial groups if, for example, exposure to enrichment activities or linguistic stimulation differs. If this is the case, then even at the same levels of care quality and exposure, children from certain racial groups may benefit more.

Racial Differences in Selection Into Preschool?

The current study focuses on the first posited explanation that racial differences in preschool effects are explained by differences in selection across groups. Several studies have modeled selection into preschool based on family characteristics (Huston, Chang, & Gennetian, 2002; Liang, Fuller, & Singer, 2000; NICHD Early Child Care Research Network, 1997). Family income is consistently a key predictor of preschool participation. Children whose parents are in the top income quartile are 23 percentage points more likely to attend early education programs than those in the bottom quartile (Bainbridge, Meyers, Tanaka, & Waldfogel, 2005). At the same time, government preschools have traditionally been targeted toward children in poverty, such that very poor children participate at higher rates than near-poor families (Meyers, Rosenbaum, Ruhm, & Waldfogel, 2004).

Many of the family characteristics that are predictive of care choice (i.e., income, maternal education and employment and family structure) differ across racial groups. For example, the ECLS–B data show that Hispanic children are much less likely than non-Hispanic White children to have working mothers or to live with a single parent, both traits associated with preschool participation. Group mean differences such as these may explain, at least in part, the differential rates of participation. Care choices may vary by race even when comparing families who are observably identical in resources.

One explanation may be omitted variables. For instance, cultural norms and beliefs about family and childrearing influence care decisions but are rarely observed by researchers (Liang et al., 2000). If these types of beliefs vary across groups, they may play a role in explaining selection differences (Fuller, Eggers-Pierola, Holloway, & Liang, 1996). In addition, child-care supply may differ by groups and non-English speakers may have trouble accessing child-care information (Liang et al., 2000).

The Current Study

In this study I use ordinary least squares (OLS) and propensity score matching to estimate the effects of preschool participation at age 4 on cognitive outcomes. I examine whether the racial differences in preschool effects reported in several recent studies are replicated in the ECLS–B. This study contributes to the existing literature in two important ways. First, because of the unique strengths of the ECLS–B data, the study allows for a more careful estimation of preschool effects than has been possible with previously used data sets. The ECLS–B data are large and nationally representative, collected before and during child-care use, and include unusually rich measures of child and family characteristics. At every wave parents are asked detailed questions about their child's current and previous child-care use. The ECLS–B therefore provides a much richer picture of children's care trajectories from birth onward than other data sets commonly used to study child care. Further, because the ECLS–B relies on prospective rather than retrospective reports of care use, there is less risk of recall error (Bound, Brown, & Mathiowetz, 2000).

The ECLS–B also provides detailed ''pretreatment'' data on children and their families. Although certain characteristics such as race or gender will not change due to care participation, others, such as children's cognitive abilities or parents' reading habits, might. Many studies estimate preschool effects using data from the years *during* or *after* the child enters care. This makes it difficult to sort out causality. For instance, are parents who read to their children often more likely to send their children to preschools, or does sending children to preschools raise parents' awareness of the importance of reading? Having data collected *before* children enter care eliminates a major hurdle in estimating causal relations.

The second way this study adds to the literature is by distinguishing racial differences in preschool effects from socioeconomic differences. This is an important point when trying to accurately gauge the impact of current preschool expansion initiatives on racial achievement gaps. In a recent study, Magnuson and Waldfogel (2005) provide back-of-the-envelope estimates of how various preschool expansion models might affect racial achievement gaps. Their calculations suggest that programs that specifically boost Hispanic and Black enrollment in preschool will be the most powerful for narrowing gaps. However, their analysis does not take into account that the effects of preschool differ, both across race and across socioeconomic groups. They might therefore underestimate the possible gains from a universal preschool system. The current study provides estimates of preschool effects across racial subgroups in order to enhance our ability to predict the impact of preschool expansions on racial achievement gaps.

Method

Participants

This study relies on data from the ECLS–B, which is the first nationally representative study to track children from birth through school entry. Data are currently available from the first three waves when the children are approximately aged 9 months, 2 years, and 4 years. The base sample includes approximately 10,700 infants born in 2001 who were selected primarily from the National Center for Health Statistics vital statistics system. Note that in compliance with National Center for Education Statistics rounding rules, all sample sizes presented in this study are rounded to the nearest 50. A cluster list-frame design was used, and primary sampling units were defined by contiguous counties. Children who died or were adopted before 9 months were excluded, as were children born to mothers younger than age 15.

A key goal of the current study is to investigate the impact of preschool at age 4 controlling for child and family characteristics at earlier waves. To accomplish this, the analytic sample is limited to children who were still in the sample at the 48-month wave. The analytic sample includes all children for whom child care and family data were collected at each wave and for whom cognitive assessments were conducted at every wave. Due to attrition, this restricted sample includes approximately 7,500 children. Children who remained in the sample and were assessed at the 48-month wave differed from those who attrited or were not tested on a host of observable characteristics. The ECLS–B provides a series of sampling weights to account for this nonresponse and to allow for population-level estimates. In all analyses presented in this study, I reweight the sample using the W3C0 weight that is designed specifically for studies that consider cognitive outcomes during the 48-month wave based on child and family measures from the previous two waves. I eliminate approximately 150 additional observations (2%) due to missing data on key covariates. Specification checks in which I impute this missing data yield almost identical results. The final analytic sample includes approximately 7,400 children.

Measures

At every wave, trained interviewers conducted home visits that included detailed parent interviews and direct child assessments. Data were collected from parents using computer-assisted personal interviewing as well as self-administered questionnaires. In almost all cases, the focal child's mother or female guardian completed the interview (100% at the 9-month wave and 95% at the 48-month wave). The interview protocol was programmed in both English and Spanish, and bilingual interviewers as well as interpreters were available to assist parents (Snow et al., 2007).

Child-care arrangements. Parent-reported child-care information was used to identify each child's primary arrangement. I constructed five mutually exclusive groups to delineate primary care type at 48 months: preschool, Head Start, nonrelative care, nonparental relative care, and parental care. Preschool is defined as any classroom arrangement including day care programs, private or public centers and state prekindergartens, but excluding Head Start. Head Start is a federally funded intervention targeted toward very low-income children. It is unique from other preschool programs both in its funding and its commitment to comprehensive services including health, nutrition, and social services. Nonrelative care encompasses arrangements with home child-care providers, nannies, babysitters, or neighbors.

I defined primary care type based on total time spent in a particular care *type*. For instance, if a child attended two separate preschools, I classified them based on the total time spent in both settings. Further, if a child spent *any* time in a nonrelative, preschool or Head Start setting, they were categorized as using that care type even if they spent more hours in parental or relative care. For the 9- and 24-month waves, there are only four possible arrangements; Head Start primarily serves 3- and 4-year-olds, and questions about Head Start were only introduced in the third wave.

Cognitive outcomes. During the third wave of data collection, cognitive assessments were conducted as close as possible to the fall of the "preschool" year. Although the bulk of children were assessed between August and December 2005, some children were assessed as late as June 2006 (Snow et al., 2007). The cognitive assessments during this "preschool" wave were designed specifically for the ECLS–B study and include items from several reliable early assessments including the PreLAS 2000, the Peabody Picture Vocabulary Test, the Preschool Comprehensive Test of Phonological and Print Processing, and the Test of Early Mathematics Ability (Duncan & DeAvila, 1998; Dunn & Dunn, 1997; Ginsburg & Baroody, 2003; Lonigan, Wagner, Torgeson, & Rashotte, 2002). Items were also taken from the Family and Child Experiences Study, the Head Start Impact Study, and the ECLS–K (Zill & Resnick, 1997). The literacy assessment includes measures of phonological awareness, letter sound knowledge, letter recognition, print conventions, and word recognition. Throughout the analysis I use standardized scores with a mean of 50 and a standard deviation of 10. These *T* scores provide norm-referenced measures of student achievement.

Cognitive assessments for the first two waves were conducted as close as possible to age 9 and 24 months, respectively. In these two earlier waves, cognitive ability was assessed using the Bayley Short Form–Research Edition (BSF–R), which was designed specifically for the ECLS–B and is made up of selected items from the Bayley Scale of Infant Development (Second Edition), an instrument commonly used by psychologists to assess the development of infants and toddlers (Bayley, 1993; Psychological Corporation, 2001). At 9 months, the BSF–R measures infants' ability to explore objects, explore objects purposefully, babble, do early problem solving and communicate with words. At 24 months, the BSF–R includes measures of memory, means-end behavior, exploratory competence, object permanence, expressive communication, and receptive communication.

A striking feature of the test scores is the emergence of cognitive gaps across racial groups between the first two waves (see Fryer and Levitt, 2006, for a discussion). The top panel of Table A1 in the Appendix shows that, at the 9-month wave, the raw achievement gap between Whites and other groups ranges from a statistically insignificant .06 of a standard deviation for Hispanic children from homes in which English is the primary language and .15 of a standard deviation for Black children. By the 24-month wave, this gap has grown to .36 for Hispanic children from homes in which English is the primary language, .53 for Black children, and .78 for Hispanic children from homes in which Spanish is the primary language. Direct comparisons of the achievement gaps at 24 and 48 months are not possible because different metrics are used to assess qualitatively different skills. Nonetheless, the cognitive gaps at the ''preschool wave'' are substantial.

Race and ethnicity. Children's race is defined based on parent reports during the 9-month wave. I report results for children whose parents identified them to be White, Black, or Hispanic. Due to small sample size, preschool impacts for other subgroups (Asians, Pacific Islanders, American Indians, and children classified as ''other'' or ''multiple'') are not reported. The Hispanic subgroup is further divided based on whether English is the primary language spoken in the home. This measure is meant to proxy for both English proficiency and acculturation. Cabrera, Shannon, West, and Brooks-Gunn (2006) argue that language ability, which is correlated with years in the United States, provides a robust measure of acculturation. Other controls for English proficiency and acculturation include: whether English is *ever* spoken in the home, if the mother indicated she speaks English ''not very well'' or ''not well at all,'' and whether the mother was born in the United States.

Socioeconomic status. Several measures are used to account for SES. The first is a standardized continuous, composite variable constructed based on parent-reported information from the 9-month wave ($M = 0$, $SD = 1$). Components of this SES index include household income, parents' employment status and parents' education level. The ECLS–B also provides a set of composite variables that use family income and size to determine whether a family is above or below various federal poverty thresholds. In this study analyses are disaggregated based on whether a family is below 130% of the poverty line. Just over one third of the sample is in this group, although there are substantial differences by race. Approximately 21% of White children in the sample are below 130% of the poverty line. The figure is 60% for the Black children, and 49% for Hispanic children. Despite the fact that most Black children in the sample are below the threshold, there are still approximately 450 Black children above this poverty threshold, providing a sufficient sample size for this subgroup analysis.

Additional covariates. A key advantage of the ECLS–B is the availability of extremely detailed information about family resources and environment. I make use of a large set of the covariates to better account for nonrandom selection into care. Descriptive statistics for these covariates, measured at the 9-month wave unless otherwise noted and disaggregated by race, are presented in the lower panel of Table A1 in the Appendix. First, I account for key child characteristics as reported on birth certificates. These include age (calculated from date of birth), gender, birth weight, multiple status, and birth order.

Next, I control for maternal employment and educational attainment as well as family structure because these factors are known to be strong predictors of care choice and also differ by race. Maternal employment is included in the model using three indicator variables (mother works full-time, mother works part-time, mother is unemployed). Mother's educational attainment is entered into the model as three indicator variables as well (mother did not graduate from high school, mother went to a vocational school or completed some college, and mother has a bachelor's degree or an advanced degree). To account for family structure, I control for whether the child is being raised by a single parent, the age of the mother when she had her first child, and the number of children and adults living in the home.

An additional set of covariates is included to more thoroughly account for differences in family resources and to capture use of targeted public programs. These include: whether the family experienced food insufficiency (medium and high food insufficiency are defined based on the frequency with which food is lacking in the home); whether the family received benefits from the Nutrition Program for Women, Infants, and Children (WIC) in the past year; whether the family received Food Stamps or welfare since the child's birth; whether the family owns a car; and whether they live in government-subsidized housing.

The ECLS–B also includes measures of parenting practices, home-learning environment, maternal stress levels, and social support. These variables are included because they capture traits that are likely to be correlated with child development and potentially are also correlated with care choice. Measures of parenting and home environment include: whether the child was breast fed, frequency of joint book reading in the home, and mother's knowledge of basic child development concepts (sum of 11 items about typical child development). Measures of community and social support include whether the mother indicated she had anyone to turn to in stressful situations (i.e., emergency, needing parenting advice, needing to borrow money) and whether she was involved in a church on a weekly basis. Differences in these measures of social support may be important for predicting differences in preschool care utilization across groups.

Finally, the models control for region of the United States (Northeast, Midwest, South) and community type (rural, suburban). These are included to account for regional differences in resources, services for children and preschool availability.

Data Analysis

I estimate OLS regressions predicting each child's literacy score at the preschool wave (roughly 48 months) as a function of their current primary care type, which is measured with a series of indicator variables. Children in parental care serve as the omitted (or base) category. Models are adjusted for each child's age at the time of the 48-month assessment, as well as a vector of controls (discussed earlier) that adjust for child and family characteristics as reported at roughly 9 months. I then interact each of the nonparental care types (preschool, Head Start, nonrelative, and relative) with each of the racial indicators to measure whether the impact of preschool varies significantly across groups. Identical models predicting mathematics ability are also explored. They yield substantively similar results and to save space are omitted.

The OLS models rely on the rich set of controls available in ECLS–B to adjust for key differences in child and family characteristics across care type. These models provide plausible estimates of preschool effects only if *all* relevant covariates are included; omitting a variable that is correlated with the outcome and predictive of preschool use will bias the results. There is no way to explicitly test whether all relevant variables are included, however, the rich set of controls available make the assumption more credible than has been possible with previously used data.

As a specification check I also replicate each analysis using a multinomial propensity score matching approach (Lechner, 2002; Robins, Hernán, & Brumback, 2000). Although the rich set of controls available in the ECLS–B goes a long way toward reducing selection bias concerns, a potential problem with the OLS models is extrapolation beyond the region of ''common support.'' To make compelling causal claims about the effects of a particular treatment, we would ideally like to compare observably similar individuals who both did and did not participate in the treatment. However, if some of the people who experienced a particular treatment are completely different from any of the people who did not receive the treatment, we cannot accurately measure the effect the treatment had for this group. Using OLS, inferences are made over the entire range of the data, even in areas, where there is no common support. Propensity score matching provides one useful solution to this problem by explicitly identifying the region of common support and eliminating observations for which no appropriate controls exist.

Matching strategies have become increasingly common for estimating treatment effects. Each individual who receives a particular ''treatment'' (i.e., preschool care) is matched to one or more individuals who are similar on observable pretreatment characteristics but did not experience the treatment. Assuming that individuals are matched on *all* factors that influence both selection decisions and child outcomes, the matched pairs simulate an experimental design. Researchers typically identify ''similar'' individuals based on their propensity score or their probability of receiving the treatment. I estimate each individual's likelihood of being in each of the five care arrangements using a multinomial logit model that includes the covariates described earlier and several higher order terms. Following cutoff suggestions presented in Caliendo and Kopeinig (2008), I then limit the sample to children for whom causal inferences can more reliably be made. I reweight the data using inverse probability of treatment weighting (IPTW; Robins et al., 2000). Each individual's observation is reweighted by the inverse of their probability of being in their actual treatment group. This reweighting process reduces selection bias by giving higher weights to children who could plausibly be in treatments other than the one in which they are observed. Without IPTW reweighting, almost all of the covariates in the model are significantly related to care choice. In other words, children in each care type during the 48-month wave differ systematically from each other on most characteristics observed at 9 months. Once reweighted, race is no longer systematically related to treatment and the bulk of covariates do not differ in significant ways by care type. A table showing these differences before and after reweighting is available from the author upon request.

Distinguishing racial differences from income differences. In the final portion of the analysis, I examine the extent to which observed differences in preschool effects across racial groups are driven by SES differences. I explore this issue by reestimating the regression model with Race × Preschool interaction terms for the subsamples below and above a poverty threshold. I also run regressions on the full sample, but add an SES × Care interaction term. The purpose of both sets of models is to assess whether allowing preschool effects to vary with SES eliminates the observed racial differences.

Results

Results are presented in three sections. First, differences in preschool use across groups are highlighted. The second section presents OLS and propensity score matching estimates of the effect of preschool participation on literacy outcomes at roughly 48 months. Both aggregate impact and differences across racial groups are considered. The final section describes results from models that allow the impact of care to interact with SES.

Differences in Care Participation

Table 1 shows child-care participation rates across waves as well as racial groups. The first column presents trends for the full sample. When the children are approximately 9 months old, about half are cared for solely by their parents and another quarter have a regular care arrangement with relatives. At this early age, less than 10% of children attend a preschool. When the children are approximately 24 months, half are still in parental care but there is a shift from relative care to preschool, with approximately 17% of children in each of these groups. Many preschool programs are targeted toward 4-year-olds, so it is not surprising that the most meaningful changes in primary care type occur between the 24- and 48-month Waves. At the 48-month wave, the percentage of children in parental and relative care drops to 18% and 7%, respectively, whereas more than two thirds are enrolled in either a preschool or Head Start program.

The trends for the full sample mask substantial differences across race. In the first two waves, Black children attend preschool at a much higher rate than other children. At the 24-month wave, for example, 26% of Black children are enrolled in preschool compared with 17% and 12% of White and Hispanic children, respectively. At the 48-month wave, the distribution of children between preschool and Head Start also varies substantially by race. Sixty percent of White children are enrolled in preschool and only 8% are enrolled in Head Start. In contrast, 32% of Black children and 21% of

Table 1
Primary Care Participation by Race and Ethnicity (9, 24, and 48 Months)

Primary care	Full sample	White	Black	Hisp	Hispanic English	Hispanic Spanish
Wave 1 ≈ 9 months						
Parents	0.493	0.513	0.352	0.530	0.495	0.564
Relatives	0.243	0.199	0.340	0.283	0.273	0.293
Nonrelative	0.171	0.190	0.166	0.138	0.157	0.120
Preschool	0.093	0.098	0.142	0.049	0.075	0.023
Wave 2 ≈ 24 months						
Parents	0.493	0.508	0.355	0.540	0.493	0.585
Relatives	0.172	0.130	0.265	0.215	0.217	0.213
Nonrelative	0.162	0.188	0.124	0.130	0.140	0.119
Preschool	0.172	0.174	0.256	0.115	0.150	0.082
Wave 3 ≈ 48 months						
Parents	0.183	0.178	0.150	0.222	0.197	0.247
Relatives	0.069	0.050	0.086	0.099	0.127	0.072
Nonrelative	0.081	0.095	0.050	0.065	0.071	0.060
Preschool	0.522	0.602	0.396	0.405	0.448	0.363
Head Start	0.145	0.076	0.318	0.209	0.157	0.259
N[a]	7,400	3,400	1,150	1,300	750	550

Note. Each column represents a subgroup, and the table is populated with the proportion of children in that subgroup who use a particular form of care at each wave.
[a] In compliance with National Center for Education Statistics rounding rules, all sample sizes are rounded to the nearest 50.

Hispanic children use Head Start as their primary care choice.

The table also highlights differences in care utilization between Hispanic subgroups. Relative to Hispanic children whose families primarily speak English in the home, Hispanic children whose parents primarily speak Spanish are more likely to be in parental care. They are also more likely to enroll in Head Start than in a preschool. The large differences in preschool utilization across racial and ethnic groups highlight the need to account for selection differences in estimating the impact of preschool participation.

Aggregate Effects of Preschool Participation

Table 2 presents OLS and propensity score matching estimates of the associations between care type and literacy outcomes, both measured during the 48-month wave. Column 1 shows estimates of the ''raw'' associations between care type and child outcomes, adjusted only for the child's age at assessment and age squared. Parental care is the omitted category. This first model highlights substantial differences in cognitive outcomes by primary care choice. Children who attend preschools score more than 4.5 points higher on the literacy assessment. Recall that the cognitive assessments are normed with a standard deviation of 10, so these coefficients amount to just under half a standard deviation. Children in nonrelative care also outperform those solely in parental care by about 2.3 points. In contrast, those in relative care perform about 1.7 points worse. Notably, in these unadjusted models Head Start attendees perform about as well as children who stay at home with their parents. Given the high levels of poverty that characterize Head Start participants, it is surprising that the unadjusted Head Start coefficients are not negative.

These ''raw'' estimates are not appropriate measures of the causal impact of care choices as they do not account for *any* child or family characteristics. In Column 2, I control for child gender and birth weight, SES, participation in WIC (a needs-based nutrition program), and mother's age at first childbirth. These covariates were selected based on Fryer and Levitt (2004), which shows that this short list of characteristics explains away the achievement gap between Whites and Blacks at kindergarten entry. Accounting for these variables reduces the association between preschool and literacy outcomes substantially, though the coefficient for preschool is still large and highly statistically significant. Not surprisingly, the Head Start coefficient increases once I account for these basic controls. In fact, the Head Start coefficient is not significantly different from the coefficients for preschool.

Column 3 shows results from models that also control for race. After accounting for the key demographic variables, race does not substantially alter

Table 2

Ordinary Least Squares and Propensity Score Estimates of the Relation Between Care Participation and Literacy Outcomes Measured During the 48-Month Wave

	1	2	3	4	5	6
Preschool	4.609*** (0.402)	1.781*** (0.378)	1.649*** (0.375)	1.316*** (0.364)	1.253** (0.424)	1.390*** (0.379)
Head Start	−0.224 (0.489)	1.544*** (0.455)	1.494** (0.457)	1.447** (0.447)	1.797** (0.667)	1.639*** (0.463)
Nonrelative	2.329*** (0.618)	−0.135 (0.541)	−0.198 (0.537)	−0.395 (0.529)	−0.260 (0.631)	−0.553 (0.576)
Relative	−1.745** (0.612)	−0.806 (0.551)	−0.895 (0.545)	−0.734 (0.526)	−1.521* (0.605)	−0.710 (0.545)
Adjusted R^2	.115	.270	.282	.320	.361	.299
N	7,400	7,400	7,400	7,400	6,600	6,600
Model specification						
Unconditional	X					
Basic controls		X	X	X	X	X
Race			X	X	X	X
Full controls				X	X	X
Region of common support					X	X
Matching (IPTW) reweighting)					X	

Note. Standard errors are in parentheses. All models include controls for age and age-squared. Basic controls (Model 2) include child gender and birth weight, socioeconomic status, participation in WIC (a needs-based nutrition program), and mother's age at first childbirth. ''Full controls'' includes all child and family covariates as discussed in the measures section and presented in Table A1 in the Appendix. In compliance with National Center for Education Statistics rounding rules, all sample sizes are rounded to the nearest 50.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

the results or add much explanatory power. In fact, the results remain stable in column 4 when I add the full set of child and family characteristics from the 9-month wave, including the child's 9-month mental and motor assessments.

Four alternative specifications were also considered, although results are not shown. First, a model that controls for children's cognitive outcomes at the 24-month wave suggests slightly higher preschool and Head Start coefficients. However, due to the substantial number of children who already attend preschool by the second wave, I worry that including cognitive measure from the 24-month wave adds endogeneity to the model. A second, more flexible model interacts race with key child and family characteristics (i.e., child's gender, maternal education, etc.). Although some of these interaction terms are significant predictors of cognitive outcomes, including them has no impact on the child-care coefficients. Finally, I consider models with state and then county fixed effects to account for unobservable, within-region characteristics that may be related to care choices as well as cognitive outcomes. Again, adding these fixed effects does not meaningfully impact the preschool or Head Start coefficients.

For simplicity and ease of interpretation, I prefer the model shown in column 4 and move forward with it for the remainder of the study. The results from this specification suggest that preschool participation is associated with literacy outcomes that are .13 of a standard deviation higher than those observed for children in parental care. The association between Head Start and literacy outcomes is quite comparable (.14 of a standard deviation). Children in nonrelative or relative care perform about as well as their peers in parental care.

Column 5 shows analogous results from the propensity score models. The first thing to note is that once the analysis is limited to the region of common support, the sample size drops from approximately 7,400 to 6,600. Eliminating observations outside the region of common support involves an inherent trade-off. The new sample includes children who are observably more similar, and therefore more appropriate for causal analysis. On the other hand, by eliminating observations I am left with a smaller sample, less precision, and potentially less generalizability. In the current analysis, the 10% of the analytic sample that lies outside the region of common support is very different from those within. They are more likely to be White or Asian and their SES is much higher. Although 25% of mothers in the region of common support have acquired a bachelor's degree or more, the figure is 70% for the omitted group. The omitted group is also about 15 percentage points more likely to be enrolled preschool (67% vs. 52%).

The results from the propensity score estimates are very similar to the original OLS estimates. The estimated effect of preschool is stable and the estimated impact of Head Start rises somewhat (from 1.4 to 1.8). This modest change is in the expected direction given that children with no appropriate matches are eliminated. Rerunning the OLS model on the same subset of children yields comparable results (see column 6). Given the extensive covariates available in the ECLS–B, it is not surprising that the two approaches provide similar estimates. Both models show that preschool participation is associated with just over one tenth of a standard deviation higher literacy outcome. The point estimates for Head Start, though not statistically different from those reported for preschool, are slightly higher.

Differences in Preschool Effects by Race

A key goal of this study is to examine whether the preschool effects shown earlier differ systematically across racial groups. To get at this, I run a model predicting literacy outcomes based on the preferred specification (Model 4 of Table 2) and a set of Race × Care interaction terms. Results from this model are presented in the top panel of Table 3. The first column shows main effects, which can be interpreted as the association between care choices and test scores for the White subgroup. The main effects for preschool and Head Start are small in magnitude and statistically insignificant. In contrast, the interaction between Black and preschool is quite large and statistically significant. Black children who attend preschool score 3.4 points higher (or an effect size of 0.34) than their peers primarily in parental care. This is the sum of the main preschool effect (0.479) and the interaction term between preschool and Black (2.905). Among Hispanics, the preschool interaction terms are statistically indistinguishable from those of Whites, although for children from primarily Spanish-speaking families the magnitude of the point estimate is large. Further, among both Hispanic subgroups, the estimated impact of participating in Head Start is substantial (effect sizes = 0.29–0.32) and higher than the insignificant effect observed among White children.

The p value on the joint F test that the Race × Preschool interactions jointly equal zero is

Table 3

Ordinary Least Squares Models Estimating the Relation Between Care Choices and Literacy Outcomes (With Race × Treatment Interactions)

	Main effects (White)	Black	Hispanic-English	Hispanic-Spanish	p values[a]
Panel 1: Full sample					
Preschool	0.479 (0.479)	2.905** (1.108)	0.956 (1.064)	1.700 (1.187)	.0037
Head Start	0.540 (0.741)	1.125 (1.251)	2.366† (1.426)	2.671† (1.363)	.3782
Nonrelative	−1.130† (0.669)	1.090 (1.688)	2.658 (1.762)	0.770 (1.748)	.6314
Relative	−1.482† (0.830)	3.221* (1.557)	0.842 (1.526)	1.458 (1.549)	.4699
p value[b]		.0515	.4012	.3715	
Panel 2: Below 130% of the poverty line					
Preschool	2.590** (0.846)	0.874 (1.461)	0.713 (1.818)	−1.810 (1.742)	.021
Head Start	1.847† (0.990)	−0.439 (1.526)	1.635 (1.801)	1.954 (1.755)	.158
Nonrelative	−0.695 (1.156)	−0.296 (2.484)	4.018 (3.355)	2.495 (2.455)	.0253
Relative	−1.294 (1.163)	3.913* (1.914)	−0.693 (2.096)	1.579 (2.043)	.3346
p value[b]		.1911	.6553	.2073	
Panel 3: Above 130% of the poverty line					
Preschool	−0.140 (0.591)	3.748† (1.967)	0.536 (1.563)	3.154† (1.651)	.1743
Head Start	−0.692 (1.103)	3.245 (2.251)	2.671 (2.355)	2.912 (2.184)	.3709
Nonrelative	−1.455† (0.806)	1.646 (2.628)	1.782 (2.209)	−1.425 (2.150)	.3617
Relative	−1.342 (1.125)	0.264 (2.833)	1.532 (2.174)	−1.119 (2.228)	.8800
p value[b]			.7611	.0463	

Note. Each panel represents a single regression of children's test scores on their care type and a set of Care × Race interactions. Models also include the full set of control covariates. Interactions between care choices and the "other" and "multiple" race groups are also included such that the main effect can be interpreted as the effect for White children. Sample size (rounded to nearest 50): Panel 1–7,400; Panel 2–2,500; Panel 3–4,900. Standard errors are in parentheses.

[a]For each care type, are the Care × Race interactions jointly significant? [b]For each race, are the Care x Race interactions jointly significant?

†$p < .10$. *$p < .05$. **$p < .01$.

easily rejected ($p = .004$). The model provides support for the claim that the effects of preschool vary systematically by race. Results from identical models that use the propensity matching approach (not shown) are similar; however, the interaction terms between Head Start and each of the racial subgroups are significant at the .1 level, and F tests imply that both the Race × Preschool and Race × Head Start interactions are jointly significant.

Taken together the results suggest differences in the impacts of early childhood interventions across groups. Black children appear to benefit more from attending preschool relative to White children, and there is some indication that the impact of Head Start may also differ across groups.

Explaining the Racial Differences: Race as a Proxy for Resources

One candidate explanation for the differences in preschool effects across groups is that race is serving as a proxy for economic resources. Previous research suggests that low-income children experience greater returns from preschool interventions relative to children whose families have more resources. This relation is also observed in the ECLS–B. Table 4 shows that children whose families have the lowest income to poverty ratios experience large and significant preschool effects. In

Table 4

Ordinary Least Squares Estimates of the Effects of Child-Care Settings on Literacy Outcomes by Income-to-Needs Ratio (Lowest Quartile, Middle Half, and Upper Quartile)

	Low	Middle	High
Preschool	3.369*** (0.685)	0.766 (0.482)	−0.786 (0.994)
Head Start	2.181** (0.678)	1.421* (0.629)	N/A
Nonrelative	0.383 (1.033)	−1.132 (0.716)	−1.507 (1.266)
Relative	0.727 (0.876)	−1.580* (0.720)	N/A
R^2	.244	.246	.246
N	1,800	3,900	1,700

Note. The income-to-needs ratio is calculated as the ratio of a families' (approximate) income to the poverty level for a family of their size. The "low" category includes the bottom quartile of families whose income-to-needs ratio falls under 1.05 (just slightly over the poverty line. The "high" category includes the top quartile of families (income-to-needs > 4.09). Poverty thresholds were taken from the ECLS–B Preschool Year User's Manual. Estimates for Head Start and relative care are omitted from the final column due to small size. They each make up less than 3% of the observations in the high income-to-needs category. Standard errors are in parentheses. In compliance with National Center for Education Statistics rounding rules, all sample sizes are rounded to the nearest 50.

*$p < .05$. **$p < .01$. ***$p < .001$.

contrast, among the middle half of the sample and the top quartile there is no association between preschool and cognitive outcomes. Because Black and Hispanic children are overrepresented in the lowest group, it may be the case that poverty status explains what appears to be differential effects by race.

The two lower panels of Table 3 present results from OLS models with Race × Care interactions estimated separately for subsamples below and above 130% of the poverty line. If differences in the impact of preschool across racial groups are driven by differences in their poverty status, Race × Care interactions will be weak predictors of cognitive outcomes in these restricted subsamples. This is, indeed, what the results from the poverty sample show. The main preschool effect, which corresponds to the impact of preschool for poor White children, is large and significant (.26 of a standard deviation) and the Race × Care interactions are negligible. This model suggests that on average all low-income children—irrespective of race—respond positively and fairly similarly to preschool participation. This is consistent with the "race as a proxy for resources" explanation.

Among families above 130% of the poverty line, the main effects are small and insignificant. For White children in this subsample, children who attend preschool perform as well as children who are primarily in parental care. If all children living above the poverty threshold responded similarly to preschool, the Race × Preschool interactions would again be insignificant. This is not the case. The results show that the impact of preschool for both Black children and Hispanic children whose families primarily speak Spanish is large and significantly greater than the impact for White children. Although I cannot reject the null hypothesis that the set of Race × Preschool interaction terms are all equal to zero, the magnitude of the coefficients suggests that among the nonpoor sample, preschool impacts may differ with race.

Table 5 shows results from similar OLS models with added interaction terms between SES and each care type. Including these interaction terms allows me to retain the full sample and is more flexible than using a relatively arbitrary poverty cutoff. The results indicate that the interaction between preschool and SES is negative and significant. This again suggests that the benefits of preschool over parental care diminish as family resources rise. Accounting for this SES × Preschool interaction, however, does not eliminate the large and significant coefficient on the Black × Preschool interaction. Further, even when accounting for the interaction between SES and preschool attendance, I can still reject the null hypothesis that the Race × Preschool interactions are jointly equal to zero ($p = .004$).

As a final check, I also run OLS regressions separately by race, and include the SES × Care interaction terms (see Table 6). These models are somewhat more flexible than the pooled model with Care × Race interactions because they allow each of the covariates in the model to operate differently across racial groups. The interaction term is negative in the full sample as well as for Whites and Hispanics. In contrast, among Black children, the SES × Care interaction is actually positive, and marginally significant ($p = .110$). Among Black children the positive effects of preschools do not appear to be inversely related with poverty.

Table 5

Ordinary Least Squares Models Estimating the Relation Between Care Choices and Literacy Outcomes at the 48-Month Wave, Including Race × Treatment and SES × Treatment Interactions

Literacy	Main effects (White)	Black	Hispanic-English	Hispanic-Spanish	SES	p values[a]
Preschool	0.554 (0.478)	1.987† (1.151)	0.464 (1.070)	0.644 (1.217)	−1.466** (0.503)	.0039
Head Start	0.445 (0.820)	0.436 (1.281)	1.993 (1.429)	1.906 (1.359)	−1.120 (0.740)	.4861
Nonrelative	−1.195† (0.694)	0.398 (1.755)	2.293 (1.754)	0.013 (1.849)	−1.112 (0.688)	.7877
Relative	−1.449† (0.861)	2.662† (1.592)	0.540 (1.530)	0.825 (1.564)	−0.739 (0.804)	.7067
p Value[b]		.2229	.5378	.6984	.0672	

Note. The panel represents a single regression of children's literacy test scores on their care type, a set of Care × Race interactions, and an SES × Care interaction term. The terms in the "SES" column are the coefficients on the interaction terms. Models also include the full set of control covariates. Interactions between care choices and the "other" and "multiple" race groups are also included (but not shown) such that the main effect can be interpreted as the effect for White children. Sample size (rounded to the nearest 50)—7,400. Standard errors are in parentheses.

[a] For each care type, are the Care × Race interactions jointly significant? [b] Joint test of significance for Treatment × Race interactions (by race).

$^{\dagger}p < .10$. $^{**}p < .01$.

Table 6

Ordinary Least Squares Models Estimating the Relation Between Care Choices and Literacy Outcomes at the 48-Month Wave, Including SES × Treatment Interaction (by Race)

48-month literacy score	Full		White		Black		Hispanic		Hispanic-English		Hispanic-Spanish	
Preschool	1.062**	(0.380)	0.667	(0.492)	6.254***	(1.440)	0.772	(0.961)	0.646	(1.181)	0.946	(2.125)
Head start	1.055†	(0.602)	0.034	(0.990)	5.392***	(1.585)	1.436	(1.385)	3.382†	(1.910)	−0.552	(2.086)
Nonrelative	−0.797	(0.544)	−1.286†	(0.718)	2.165	(2.059)	−0.941	(1.319)	−0.004	(1.752)	−3.665	(2.575)
Relative	−1.085†	(0.637)	−1.320	(0.871)	3.906†	(1.993)	−1.962	(1.456)	−0.619	(1.881)	−3.634	(2.300)
Preschool × SES	−1.754***	(0.477)	−2.103**	(0.651)	2.232	(1.398)	−1.759	(1.232)	−0.914	(1.509)	−2.221	(2.642)
SES	4.267***	(0.615)	3.882***	(0.835)	−2.252	(1.836)	5.545***	(1.565)	5.224**	(1.973)	6.227*	(2.803)
Adjusted R^2	.311		.302		.305		.278		.341		.210	
N	7,400		3,400		1,150		1,300		750		550	

Note. Models also include the full set of control covariates. Standard errors are in parentheses. In compliance with National Center for Education Statistics rounding rules, all sample sizes are rounded to the nearest 50.
$^{\dagger}p < .1$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

These analyses provide mixed evidence for the hypothesis that racial differences in preschool effects are capturing differences in SES. The models show that, on average, all poor children, irrespective of race, benefit from preschool participation relative to parental care. At the same time, after accounting for the interaction between family resources and preschool effects, Black children still seem to experience systematically higher returns to preschool participation. Further, their larger gains from preschool seem to be driven largely by the nonpoor sample.

Discussion

This study uses a large, nationally representative data set to investigate whether the impacts of attending preschool at age 4 vary across racial groups. States and communities are increasingly considering or implementing universal rather than targeted preschool interventions. In this context, estimating the differential effects of preschool is essential for understanding whether early interventions can realistically narrow racial achievement gaps.

Using OLS and propensity score matching, I find evidence that preschool effects vary both across racial groups and between Hispanic subgroups even after controlling for SES and an extensive list of child and family characteristics. For White children, the association between preschool participation and literacy scores is small and insignificant. In contrast, effect sizes are substantial and significant among Black and Hispanic children. For Black children, the estimated preschool effect size is 0.34, which is about 3 times larger than the preschool effects reported at kindergarten entry using the ECLS–K (Loeb et al., 2007). Both Hispanic subgroups appear to benefit from preschool participation, although the estimated effect sizes are larger for Hispanic children from primarily Spanish-speaking homes (0.22) than for those whose families primarily speak English (0.14). Further, there is evidence that, among Hispanic children, the effects of Head Start are also large, a finding that is consistent with the ECLS–K results.

Although these first set of findings stem from models that statistically adjust for a rich set of covariates, the models do not explicitly account for the possibility that the relation between preschool participation and child outcomes might vary with family resources. To get at this issue I rerun the analyses for poor and nonpoor subgroups and, more flexibly with interactions between each care choice and SES. The purpose of these analyses is to assess whether the differential effects of preschool across groups diminish once I allow the impact of preschool programs to vary with SES. Indeed, I find that among very low-income children, defined as those whose family income places them under 130% of the poverty line, racial differences in preschool effects are far less pronounced, and statistically negligible. On average, poor children (irrespective of race) perform about a quarter of a standard deviation higher on a literacy assessment than those children who are primarily in parental care.

Among the nonpoor sample, however, the returns to attending preschool (relative to staying home with a parent) are much larger for Black children and Hispanic children whose families primarily speak Spanish, than for their peers. Consistent with findings from previous studies, the results also indicate that among White and Hispanic children the effects of preschool are inversely related to SES, such that the benefits of preschool are largest among the poorest children. In contrast, among Black children the interaction between SES and preschool participation is actually positive, and marginally statistically significant ($p = .110$). These results are surprising and warrant further examination.

As part of the growing debate on universal preschool, researchers and policymakers are interested in whether preschool yields meaningful positive impacts for middle-income families (Barnett, Brown, & Shore, 2004). Several studies have highlighted positive short-term benefits from preschool attendance for nonpoor children (Gormley & Phillips, 2005; Loeb et al., 2007). The current study, which finds positive preschool effects for nonpoor Black children, provides some support for this finding. However, to the best of my knowledge, it is the first study to suggest that the benefits of preschool among nonpoor children may be more pronounced for some racial and ethnic groups.

Why might nonpoor Black children benefit more from preschool participation than their peers? In this study I posited several possible explanations for differences in preschool effects across race, and I revisit them now, in light of the current findings. One possibility is that nonpoor Black children have greater exposure to preschool than their White or Hispanic peers and that this more intense preschool experience might explain the positive gains. Data from the ECLS–B, as well as other studies, show that, on average, Black children start preschool at an earlier age, and attend for more hours per week (Magnuson & Waldfogel, 2005; Snyder & Hoffman, 2002). Though previous studies have not reported whether these patterns persist among nonpoor families, if they do then the differences in preschool effects might capture the cumulative benefits of more time spent in preschool.

Another possibility is that the nonpoor Black children have systematically different preschool experiences with respect to quality, content, or focus. A few studies find that on average Black children attend preschool programs that are of lower quality than those used by White children. However, these quality differences would suggest that—if anything—gains from preschool would be smaller for Black children than for White children. The quality explanation is therefore inconsistent with the findings from this study. Still, the few studies that examine differences in preschool quality across racial groups do not disaggregate results by family resources. It might be that among nonpoor families quality differences operate in the opposite direction. Perhaps nonpoor Black children are more likely than their peers to attend preschool programs that focus on the types of preliteracy skills that are assessed in the ECLS–B.

Finally, it might be that the observed differences in preschool effects across groups stem from differences in the activities and learning environments that children in the comparison group (those in parental care) experience. More research is necessary to better understand the home-learning environments of 4-year-old children who are primarily in parental care, and to see how experiences may vary across groups.

Although an empirical exploration of the various mechanisms that explain differences in preschool impacts across groups is beyond the scope of this current study, the ECLS–B does provide a unique opportunity to test some of the hypotheses posited. At each round of data collection, the ECLS–B asks parents about their child-care arrangements, and includes items that address age at entry and hours of participation. During the 24- and 48-month waves, the focal child's preschool teacher and preschool director were surveyed. These surveys provide extensive data on the structural characteristics of the child's preschool program as well as the curriculum and the teacher's beliefs. Finally, the ECLS–B asks parents many questions about the frequency with which various learning activities occur in the home (i.e., joint book reading, trips to the library, computer usage). In future analyses, these data can be used to better understand whether intensity of preschool exposure, preschool quality, and home-learning environment partially explain differences in preschool effects across groups.

Limitations and Remaining Questions

Although the current study provides important new insights about the differential impact of preschool participation across groups, the study has several limitations worth noting. First, the study does not account for variation in the amount of preschool children have experienced at the time of their 48-month wave assessment. Variation in exposure to preschool occurs for two reasons. First, some variation occurs because the data collection for the third wave of the ECLS–B was collected over the entire preschool year. Although most children were assessed in the fall, a substantial portion was assessed later in the year. Children who were assessed later are likely to have experienced greater amounts of preschool. Preschool exposure also varies across children due to true differences in the age at which children start preschool and the hours they participate each week. For example, among children who use preschool as their primary form of child care, the average age at first preschool experience is 32 months with a standard deviation of 17 months. Children attend preschool for an average of 23 hr a week, with a standard deviation of 14 hr. The current study presents average preschool effects without accounting for the possibility that preschool impact varies by differences in exposure.

The ECLS–B provides better data on children's child-care histories than any other nationally representative dataset, including detailed information about multiple arrangements at three time points. However, the data cannot be used to construct a complete picture of each child's full exposure to various types of child care from birth to age 4. For example, if during the 48-month wave a parent indicates that their child is not currently enrolled in preschool, there is no way to discern whether the child was in preschool at *any* point since the 24-month wave data collection. A related issue is that a child who attends preschool for the first time at 48 months, and participates for 30 hr a week, will be indistinguishable from one who has been attending for 30 hr starting just after their 24-month assessment. It is therefore not possible to fully explore the issue of preschool exposure using this dataset, although analyses using somewhat cruder measures of exposure (i.e., attended preschool for at least 2 years) are certainly possible.

A second limitation of the study is that it considers only the very short-term impacts of attending preschool. Public investment in early childhood interventions is motivated by the desire to make a meaningful and lasting impact on the life of participants throughout their school years and beyond. It remains to be seen whether the positive effects observed in this study persist as the ECLS–B children continue into elementary school. Data from the final waves of the ECLS–B are not yet released. When they are, it will be possible to examine the link between preschool participation and cognitive outcomes at kindergarten entry.

Unfortunately, the design of the ECLS–B does not allow for tracking of children after school entrance. However, previous studies have shown that the impacts of preschool interventions often dissipate as children progress through school (Karoly et al., 1998; Magnuson et al., 2007). Further, there is some evidence that the extent of this fade out, at least in the case of Head Start, is systematically different for Black and White children. Currie and Thomas (2000) show that although the positive effects of Head Start among Black children disappear rapidly, White children experience long-lasting benefits through adolescence. They attribute these differences to differences in the average quality of schools that Black and White children attend. Ferguson (1998) hypothesizes that universal access to preschool might actually exacerbate gaps between White and Black children if both groups benefit, but the effects dissipate more rapidly among Black children. Understanding the extent to which positive gains from preschool fade and the extent to which this fade-out varies across groups is essential for better understanding the potential of preschool interventions to narrow long-term achievement gaps between groups.

Concluding Remarks

The current study evaluates the impact of preschool participation at age 4 using a large, nationally representative data set and controls for a detailed set of pretreatment characteristics including cognitive measures. The findings add to our understanding of the extent to which preschool interventions could meaningfully narrow socioeconomic and racial achievement gaps. The results echo those from previous studies that find that preschool participation is particularly beneficial for poor children. They demonstrate that, on average, poor children, irrespective of racial and ethnic background, benefit substantially and at about the same rate from preschool participation. The findings also underscore the importance of distinguishing racial differences in preschool effects from socioeconomic differences. They suggest that both poor and nonpoor Black children benefit significantly from preschool participation.

Between 2002 and 2008, the percentage of 4-year-olds enrolled in state preschool programs has grown from 14% to 24% (Barnett et al., 2008). Three states currently provide universal access to preschool, and others states are considering such initiatives. Expanded and universally available preschool programs are likely to narrow achievement gaps only if some subgroups experience larger benefits than others. This study provides encouraging preliminary evidence that both socioeconomic achievement gaps and Black–White achievement gaps could narrow within the context of a universal preschool program.

References

Bainbridge, J., Meyers, M. K., Tanaka, S., & Waldfogel, J. (2005). Who gets an early education? Family income and the enrollment of three-to five-year-olds from 1968 to 2000 *Social Science Quarterly*, *86*, 724–745.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5*, 25–50.

Barnett, W. S., Brown, K., & Shore, R. (2004). *The universal vs. targeted debate: Should the United States have preschool for all?* New Brunswick, NJ: National Institute for Early Education.

Barnett, W. S., Epstein, D. J., Friedman, A., Boyd, J. S., & Hustedt, J. (2008). *The state of preschool, 2008*. New Brunswick, NJ: National Institute for Early Education Research.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost benefit analysis using data from the age-40 followup. *Journal of Human Resources*, *41*, 162–190.

Bound, J., Brown, C., & Mathiowetz, N. (2000). Measurement error in survey data. In J. J. Heckman & E. Learner (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Amsterdam: North-Holland/Elsevier Science.

Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly*, *18*, 401–426.

Cabrera, N. J., Shannon, J. D., West, J., & Brooks-Gunn, J. (2006). Parental interactions with Latino infants: Variation by country of origin and English proficiency. *Child Development*, *77*, 1190–1207.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22, 31–72.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, *6*, 42–57.

Crosnoe, R. (2007). Early child care and the school readiness of children from Mexican immigrant families. *International Migration Review*, *41*, 152–181.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, *15*, 213–238.

Currie, J., & Thomas, D. (1999). Does head start help Hispanic children? *Journal of Public Economics*, *74*, 235–262.

Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of Head Start. *Journal of Human Resources*, *35*, 755–774.

Duncan, S., & DeAvila, E. (1998). *PreLAS® 2000*. Monterey, CA: CTB/McGraw-Hill.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test Third Edition (PPVTIII)*. Circle Pines, MN: American Guidance Service.

Ferguson, R. F. (1998). Can schools narrow the black-white test score gap. In C. Jencks & D. Phillips (Eds.), *The Black-White test score gap* (pp. 318–374). Washington, DC: The Brookings Institute.

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics*, *86*, 447–464.

Fryer, R. G., & Levitt, S. D. (2006). Testing for racial differences in the mental ability of young children. National Bureau of Economic Research Working Paper Series, No. 12066. Retrieved August 29, 2008, from http://www.nber.org/papers/w12066

Fuller, B., Eggers-Pierola, C., Holloway, S., & Liang, X. (1996). Rich culture, poor markets: Why do Latino parents forgo preschooling. *The Teachers College Record*, *97*, 400–418.

Garces, E., Thomas, D., & Currie, J. (2000). Longer-term effects of Head Start. *American Economic Review*, *92*, 999–1012.

Ginsburg, H., & Baroody, A. (2003). *Test of Early Mathematics Ability* (3rd ed.). Austin, TX: PRO-ED.

Gormley, W. T. (2008). The effects of Oklahoma's pre-K program on Hispanic children. *Social Science Quarterly*, *89*, 916–936.

Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, *40*, 533–558.

Gormley, W. T., & Phillips, D. (2005). The effects of universal pre-K in Oklahoma: Research highlights and policy implications. *Policy Studies Journal*, *33*, 65–82.

Haskins, R., & Rouse, C. (2005). *Closing achievement gaps* (*The Future of Children* Policy brief). Brookings Institution.

Huston, A. C., Chang, Y. E., & Gennetian, L. (2002). Family and individual predictors of child care use by low income families with young children. *Early Childhood Research Quarterly*, *17*, 441–469.

Karoly, L., Ghosh-Dastidar, B., Zellman, G., Perlman, M., & Fernyhough, L. (2008). *Nature and quality of early care and education for California's preschool-age children: Results from the California Preschool Study*. Santa Monica, CA: RAND.

Karoly, L., Greenwood, P., Everingham, S., Hoube, J., Kilburn, R., Rydell, P., et al. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*. Santa Monica, CA: RAND.

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics*, *84*, 205–220.

Liang, X., Fuller, B., & Singer, J. D. (2000). Ethnic differences in child care selection: The influence of family structure, parental practices, and home language. *Early Childhood Research Quarterly*, *15*, 357–384.

Loeb, S., & Bassok, D. (2008). Early childhood and the achievement gap. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research on education finance and policy* (pp. 517–534). New York: Routledge.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66.

Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *75*, 47–65.

Lonigan, C., Wagner, R., Torgeson, J., & Rashotte, C. (2002). *Preschool Comprehensive Test of Phonological & Print Processing (Pre-CTOPPP)*. Austin, TX: PRO-ED.

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, *122*, 159–208.

Magnuson, K., Lahaie, C., & Waldfogel, J. (2006). Preschool and school readiness of children of immigrants. *Social Science Quarterly*, *87*(Suppl.), 1241–1262.

Magnuson, K., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51.

Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children*, *15*, 169–196.

Meyers, M. K., Rosenbaum, D., Ruhm, C., & Waldfogel, J. (2004). Inequality in early childhood education and care: What do we know? In K. Neckerman (Ed.), *Social inequality* (pp. 223–269). New York: Russell Sage Foundation.

NICHD Early Child Care Research Network (1997). Familial factors associated with the characteristics of nonmaternal care for infants. *Journal of Marriage and the Family*, *59*, 389–408.

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD study of early child care and youth development. *American Psychologist*, *61*, 99–116.

Psychological Corporation (2001). *Bayley Short Form–Research Edition*. San Antonio, TX: Author.

Reardon, S. F., & Robinson, J. P. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). Mahwah, NJ: Erlbaum.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and Juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, *285*, 2339–2346.

Robins, J. M., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*, 550–560.

Rooney, P., Hussar, W., Planty, M., Choy, S., Hampden-Thompson, G., Provasnik, S., et al. (2006). *The condition of education 2006*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Rumberger, R. W., & Tran, L. (2006). *Preschool participation and the cognitive and social development of language minority students*. Santa Barbara, CA: UC Linguistic Minority Research Institute.

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40*. Ypsilanti: High Scope Press.

Snow, K., Thalji, L., Derecho, A., Wheeless, S., Lennon, J., Kinsey, S., et al. (2007). *Early Childhood Longitudinal Study (ECLS–B), Preschool Year. Data file user's manual (2005–06)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Snyder, T. D., & Hoffman, C. M. (2002). *Digest of education statistics, 2002* (Table 44). Washington, DC: National Center for Education Statistics. Retrieved September 2, 2010, from http://nces.ed.gov/programs/digest/d02/dt044.asp

Zill, N., & Resnick, G. (1997). *FACES color names and counting test*. Rockville, MD: Westat.

Appendix

Table A1

Descriptive Statistics for Selected Variables (by race)

Variable	Full sample	White	Black	Hispanic (English)	Hispanic (Spanish)
Panel 1: Assessment scores, by waves					
9-Month Mental Score	50.252	50.754	49.271	50.150	49.462
9-Month Motor Score	50.224	49.801	52.385	50.155	48.826
24-Month Mental Score	50.572	52.826	47.536	49.234	45.053
24-Month Motor Score	50.255	50.424	51.649	49.745	48.563
48-Month Literacy	50.067	51.459	48.317	48.306	45.632
48-Month Math	50.225	51.822	47.368	48.414	46.929
Panel 2: Child and family characteristics					
Age (at 48-month wave)	52.555	52.224	52.333	53.094	53.898
Female	0.494	0.495	0.482	0.493	0.493
Mom not in labor force	0.392	0.405	0.268	0.397	0.478

Table A1
(*Continued*)

Variable	Full sample	White	Black	Hispanic (English)	Hispanic (Spanish)
Mom employment full-time	0.333	0.312	0.432	0.349	0.279
Mom employment part-time	0.204	0.241	0.141	0.179	0.136
Mom unemployed	0.080	0.051	0.169	0.085	0.108
Mom < high school	0.173	0.096	0.249	0.232	0.416
Mom high school	0.294	0.263	0.370	0.338	0.332
Mom some college	0.277	0.293	0.284	0.294	0.176
Mom BA or more	0.256	0.349	0.098	0.136	0.075
Single parent	0.193	0.094	0.574	0.250	0.154
No. children in household	2.133	2.008	2.455	2.208	2.319
No. adults in household	2.169	2.119	1.977	2.226	2.525
SES	−0.038	0.210	−0.481	−0.220	−0.616
WIC benefits past year?	0.508	0.351	0.816	0.613	0.803
Food stamps?	0.201	0.131	0.466	0.240	0.177
Mom not U.S. native	0.166	0.041	0.103	0.142	0.754
Very low birth weight	0.012	0.009	0.023	0.012	0.009
Mom's age when first child born	23.905	25.218	21.093	22.365	22.161
Weekly church attendance	0.282	0.291	0.308	0.229	0.266
Read child books every day	0.328	0.399	0.233	0.273	0.169
Parent knowledge of child development (0–11, higher = better knowledge)	6.873	7.637	5.836	6.593	4.875
Northeast	0.175	0.183	0.168	0.135	0.188
Midwest	0.227	0.281	0.199	0.114	0.099
South	0.365	0.354	0.555	0.334	0.284
Suburban	0.121	0.146	0.076	0.119	0.055
Rural	0.148	0.211	0.087	0.057	0.019
Observations	7,400	3,400	1,150	750	550

Note. All variables refer to characteristics from the 9-month wave unless otherwise noted. In compliance with NCES rounding rules, all sample sizes are rounded to the nearest 50. SES = socioeconomic status; WIC = Women, Infants, and Children.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1846–1858

Movement Planning Reflects Skill Level and Age Changes in Toddlers

Yu-ping Chen
Georgia State University

Rachel Keen
University of Massachusetts

Kerstin Rosander and
Claes von Hofsten
Uppsala University

Kinematic measures of children's reaching were found to reflect stable differences in skill level for planning for future actions. Thirty-five toddlers (18–21 months) were engaged in building block towers (precise task) and in placing blocks into an open container (imprecise task). Sixteen children were retested on the same tasks a year later. Longer deceleration as the hand approached the block for pickup was found in the tower task compared with the imprecise task, indicating planning for the second movement. More skillful toddlers who could build high towers had a longer deceleration phase when placing blocks on the tower than toddlers who built low towers. Kinematic differences between the groups remained a year later when all children could build high towers.

Task demands affect the kinematics of reaching for objects in adults (Marteniuk, MacKenzie, Jeannerod, Athenes, & Dugas, 1987). Reaching slows when the hand approaches an object that will subsequently be picked up and used in a precise way compared with an imprecise way. When young adults were asked to grasp a disk (4 cm in diameter) and then to either fit it into a small well (4.1 cm in diameter) or throw the same disk inside a box (20 cm × 40 cm × 15 cm), the approach movement to pick up the disk during the fitting task was slower, with a longer duration and a longer time spent in the deceleration phase, compared with the throwing task (Marteniuk et al., 1987).

Similar findings have been observed in the healthy elderly (Weir, MacDonald, Mallat, Leavitt, & Roy, 1998) and to some extent in healthy 10.5-month-old infants (Claxton, Keen, & McCarty, 2003). Active healthy elderly slowed the approach phase of their reaching for the object and spent a longer duration in deceleration while completing the fitting task, compared with the throwing task. For infants only two measures, peak speed and average speed, distinguished between the tasks. Like adults, infants reached to pick up a ball more slowly if they were going to fit it into a plastic tube as opposed to subsequently throwing it. Claxton et al. (2003) interpreted their findings as evidence of motor planning for an upcoming task. These studies demonstrate that task demands of the subsequent action affect the kinematics of reaching for objects over a wide age range. Slowness in the approach kinematics has been assumed to reflect the person's intention to engage in a future precision action. This is striking because there was no obvious need to vary the approach with regard to the subsequent movement because the object was the same for both actions.

As would be expected, reaching kinematics differed during the action of completing a precision task, as well as during the approach phase. When young adults were asked to point at a target using their tip of index finger compared with using any part of their hand, reaching was characterized by lower peak velocity and a longer deceleration phase (Wu, Lin, Lin, Chang, & Chen, 2005). Bryden and Roy (1999) asked adults to place a peg into a hole. They varied the diameter of the peg while keeping the peg hole the same so that sometimes it fit snug-

---

This research was supported by National Institutes of Health Grant R37 HD027714 to R. Keen and by Swedish Research Council Grants 2004-12172-22115-26 and EU-5010-CONTACT to C. von Hofsten. We are grateful to Roberta Collard for loaning us the Gesell blocks. We thank Alison Sherman, Brigid Looney, Daniel Johnson, and Dana Alhaffar for help in data collection and scoring the video data.

Correspondence concerning this article should be addressed to Rachel Keen, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904. Electronic mail may be sent to rachelkeen@virginia.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0017

gly and sometimes loosely in order to examine the effect of precision on reaching kinematics. Bryden and Roy also did the reverse: varying the peg hole but keeping the peg the same size. As expected, when the precision requirement increased with either manipulation, kinematic patterns during the action had longer movement time, lower peak velocity, and a longer deceleration phase. Thus, precision demands appear to exert a similar effect on both the approach phase and the manipulation phase.

To our knowledge, no study has examined whether task demands affect reaching kinematics in toddlers. Because toddlers have more reaching experience and better motor control than infants under a year of age, their reaching kinematics should more closely resemble those of adults. We specifically chose tower building as the precision task in this study as it is an activity most of the toddlers enjoy and it requires precision when placing one block on the top of another. Age-related norms of children's ability to construct towers are well established in many standardized assessment tools, such as the Bayley Scale of Infant Development (Bayley, 1969), or the Gesell Developmental Observation (Gesell & Amatruda, 1954). Toddlers of the same age have different levels of performance in tower building. For example, in the Bayley Scale, on average, the 13.8-month-olds (range = 10–19 months) can build a two-level cube tower, 16.7-month-olds (range = 13–21 months) can build a three-level cube tower, and 23-month-olds (range = 17–30 months) can build a six-level cube tower. We know of no study that has reported whether there are any differences in reaching kinematics between the skilled and the less skilled tower-building toddlers, but one would predict that more skilled children would show more mature patterns of movement.

The adult-learning literature describes how different skill levels affect the kinematics of movements in adults (Gentiner, 1983; Larochelle, 1983). The coordination patterns between skilled and less skilled performers were distinguishable in adults (Gentiner, 1983; Larochelle, 1983; Temprado, Della-Grasta, Farrell, & Laurent, 1997). It is difficult to make predictions for young children based on adult performance, but we can hypothesize that differences in the skill with which toddlers can build block towers will be reflected in their reaching kinematics for the precision task. On the other hand, when engaged in imprecise tasks with those same blocks, movement kinematics would not be expected to differ with skill level.

In this study we examined how task demands, skill level, and object size affect the reaching kinematics of a group of typically developing toddlers. Tower building, which required precision, was compared with a "clean-up" task of putting blocks into a large container. In Study 1, we tested within the ages of 18–21 months indicated by the Gesell and Bayley Scale norms to encompass a range of ability in tower building. We hypothesized that all toddlers would (a) slow down their reaching to pick up and move blocks when building a block tower compared with the imprecise task and (b) have different reaching kinematics when manipulating large and small blocks because it is harder to build a tower with small blocks. Furthermore, in order to examine the control strategies of toddlers who were better at the precise task, we compared high- and low-tower builders (a) when building towers and placing blocks in a container and (b) when placing blocks low in the tower versus high in the tower when more care was required. We hypothesized that toddlers who could build a high tower would slow down more in the precision task but not differ on the imprecise task. We also hypothesized that their reaching kinematics would adapt to task difficulty, slowing still further as they placed the final blocks on a tall tower. Finally, in Study 2, we brought back a subset of children a year later to see how their developing motor skills affected reaching kinematics while performing those same tasks. The continuity of skill level in high- and low-tower children was also assessed. Because all children could build a tall tower at the older age, the question was whether any initial kinematic differences would remain.

General Method

Stimuli and Apparatus

The stimuli consisted of ten 2.5-in. wooden blocks, ten 1-in. wooden cubes from the Gesell Scale of Infant Development (Gesell & Amatruda, 1954), and two large open containers with sizes appropriate to block size (30 cm × 21.5 cm × 8.5 cm and 18 cm × 12 cm × 7 cm). Gesell cubes were used rather than Bayley Scale cubes because the former are wooden and the latter are plastic. Because the large blocks were wooden, the small blocks needed to be made of the same material for comparison. The containers were available to children for the imprecise task of placing or throwing blocks into them. (Note: Block size did not systematically affect kinematics or behavior and will not be considered further.)

Hand movements were recorded at 100-Hz sampling rate using two linked banks of cameras from Phoenix System (Phoenix Co., Burnaby, BC, Canada). The two camera banks were placed on either side of the child to record movements of each hand. The camera banks were positioned at a height of about 2 m from the floor and 2 m away from the child. Prior to data collection, the testing area was calibrated by moving a rod with two sensors supplied by the Phoenix Systems through out the testing area. The average root mean square error in the calibration was always < 0.9 mm for each data collection session. One digital camera (Sony, DCR-TRV510, Sony Corporation of USA, New York, NY) with a superimposed timer was placed 45° from the horizontal plane of the toddlers' left-hand side to record behavioral movements. The kinematic data from the Phoenix Systems and video recording were synchronized by time codes placed on both devices during data collection.

Procedure

After obtaining signed informed consent from the parents, toddlers were seated on the lap of their parents at a card table, with the surface around waist height of the toddlers. The experimenter was seated across the table to present the testing stimuli and encourage toddlers to engage in the activities. Two small infrared-emitting sensors (5 mm in diameter) embedded in a Velcro wrist band were placed around each of the toddler wrists to record kinematic data for the Phoenix Systems.

Two tasks (tower building and placing blocks into an open container) and two block sizes (large and small) were presented. The testing sequence was counterbalanced between large and small blocks so that half of the toddlers were presented with large blocks first and the other half with small blocks first. The session always started with the imprecise task to reduce the possible frustration of not being able to build the block tower. At the beginning of each task, the experimenter demonstrated throwing the blocks or the tower-building movements one or two times to encourage toddlers' engagement. For the imprecise task, toddlers were encouraged to grasp a block placed at their right-hand side and then place or throw it into the large container placed at their left-hand side. For the tower-building task, toddlers were encouraged to grasp the block placed at their right-hand side and to build a block tower at their left-hand side as tall as possible. The experimenter always put the first block of the tower on the toddler's left-hand side to mark the tower location. The distance between the blocks to be picked up and the container or the tower was about 20 cm; however, the starting position of the toddler's hand was not constrained. Unlike adult studies in which starting position of the hand is specified, the unconstrained starting position of toddler reaches results in a variable distance between block and hand over trials and children. If the toddler refused to perform the imprecise task, the experimenter moved on to the tower task and later went back to the imprecise task in an attempt to get a sufficient number of usable reaches. Each condition was repeated at least four times or until the toddlers were no longer interested. (See Figure 1 for photograph of child engaged in the tower-building task.)

Data Reduction

Position data were converted into three-dimensional coordinate data using Phoenix system software. All reaches were viewed and scored from the videotapes to determine the beginning and end of a reach. Two types of reaches were scored: the approach phase and the placement phase.

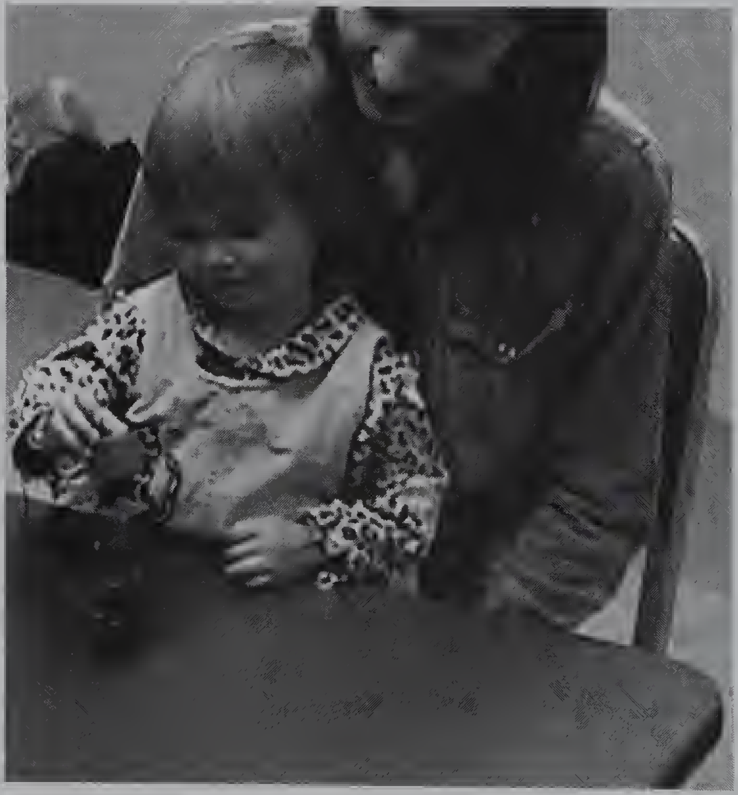


Figure 1. Photograph of a child engaged in the tower-building task using small blocks.

Note. Two small infrared-emitting sensors embedded in a Velcro wrist band were placed around each of the toddler's wrists.

The approach phase began with the first frame that the hand started to move toward the block and ended when the hand contacted the block. The placement phase started with the first frame that the block in hand started to be moved and ended at the first frame when fingers opened to release the block either into the container or on top of the block tower.

A primary observer scored all reaches from the videotape for the onset and end time of the approaching and the placement phases, as well as the tallest level of the tower each toddler could perform during the testing. These times demarcated the Phoenix data so that kinematic measures could be derived from each reach using MATLAB software (Mathworks, Natick, MA). A second observer scored all 310 reaches in 4 toddlers to calculate reliability for onset time of the approach phase (94% agreement), end time of the approach phase (97% agreement), onset time of the placement phase (95% agreement), and end time of the placement phase (78% agreement). Whenever the scored time differed by more than 0.099 s (three frames), the primary and secondary observers met to obtain agreement on the final times used for analyses ($n = 9$ for onset time of the approach phase, $n = 4$ for end time of the approach phase, $n = 9$ for onset time of the placement phase, and $n = 37$ for end time of the placement phase).

Six kinematic measures were derived for the approach and placement phases separately, including movement time (MT, defined as the time from onset to end time of each phase), straightness ratio (SR, defined as the total path length of the hand divided by the shortest distance between hand positions at the start and end frames of the reach), average speed of the reach (MEANV), amplitude of the peak velocity (PV), percentage of movement time to peak velocity (PPV, defined as the time between the start frame to the frame of peak velocity, divided by the movement time), speed at the time of contact (for the approach phase), or release of the block (for the placement phase; ENDV, defined as the speed at the end of the reach).

Toddlers' skill level of tower building was determined by the tallest level of towers they built using the 1-in. blocks from the Gesell Scale. According to the norms for this scale, the mean age of toddlers to build a two-block tower is 12 months, the mean age to build a three-block tower is 18 months, and the mean age of toddlers to build a seven-block tower is 24.0 months. (These norms are comparable with the Bayley Scale norms, which have a mean age of 13.8 months for two-block towers, 16.7 months three-block towers, and 23 months for six-block towers.) We therefore divided our toddlers into three skill levels: (a) high-tower builders who could build a four-block or higher tower, (b) low-tower builders who could build a two- or three-block tower, and (c) no-tower builders who could not build any tower. The reliability for scoring the tallest level of tower between the primary and the second observers were 100%.

For high-tower builders, we used the same criteria of determining toddlers' skill level of tower-building to determine the difficulty of reaches. Reaches during the act of tower building were categorized into difficult reaches (i.e., placing Block 4 or higher) and easy reaches (i.e., placing Block 2 or 3).

Study 1

Participants

Thirty-nine toddlers aged between 18 and 21 months participated in this study. Three were excluded from the analysis because of fussiness and refusing to wear the sensors placed on the hands, and 1 was excluded because of disinterest in the block activities, leaving 35 toddlers (19 males, 16 females; mean age = 19.7 months, *SD* = 1.2 months). The sample was composed of 35 Caucasian, middle-class children. The parents were first contacted by an informational letter describing the study, which was followed by a telephone call to see if they were interested in participating. Each toddler received a T-shirt with the lab logo and a certificate at the end of the testing as our appreciation of participation. All parents or legal guardians of the toddlers signed the informed consents in accordance with the policies of Institutional Review Board of the University of Massachusetts prior to testing.

Results and Discussion

Four children in this study did not build a tower in the small block condition, so their data were excluded from the data analyses, leaving 31 children who provided complete data for reaches in both phases of both tasks. Of those, 15 were in the low-tower group (two- or three-block tower) and 16 were in the high-tower group (four-block tower or higher).

In preliminary analyses, comparisons of the initial distance of the approach and the placement phase revealed a difference between the imprecise and tower tasks. For the approach phase, toddlers' reaches started significantly farther away from the block to be picked up in the imprecise tasks, $F(1, 22) = 13.22$, $p = .001$. The mean distance was 195 mm ($SE = 7$ mm) for the imprecise task and 161 mm ($SE = 5$ mm) for the tower task. For the placement phase, toddlers' reaches also started significantly farther away from the final release point in the imprecise tasks, $F(1, 27) = 25.32$, $p < .001$. The mean distance was 198 mm ($SE = 6$ mm) for the imprecise task and 166 mm ($SE = 4$ mm) for the tower task. Although the experimenter controlled the distance between the blocks and the container or tower, it was impossible to control the starting position and movements of the toddlers' hands. Because the distance the hand traveled could affect kinematic measures, such as MT and PV, we statistically controlled for initial distance as a covariate using mixed regression models to test the differences between tasks, skill level, and block sizes on each kinematic measure using the SAS program version 8.0 (SAS Institute Inc., Cary, NC).

We also examined gender differences for each kinematic measure during approach and placement phases. None of the kinematic measures revealed statistical differences between male and female toddlers (all $ps > .05$).

As noted earlier, the testing session in this study always started with the imprecise task, but 9 of the 31 children who built at least a two-block tower started with the tower task because they initially refused the imprecise task. An analysis of variance (ANOVA) was carried out with order as a main effect. There were no differences in the approach and placement phases for these 9 toddlers who performed the tower task first compared to others who performed the imprecise task first.

In the analyses comparing skill level of groups, only reaches for the approach and placement of Blocks 2 and 3 were used because the low-tower group, by definition, did not contribute reaches beyond this level. The average number of reaches per child during the approach phase was 6.23 ($SD = 5.43$) for the tower task and 10.09 ($SD = 7.62$) for the imprecise task. The average number of reaches per child during the placement phase was 10.49 ($SD = 6.53$) for the tower task and 13.20 ($SD = 8.09$) for the imprecise task. Reaches to blocks placed higher in the stack were analyzed separately for the high-tower group. The data for each kinematic measure were analyzed with a mixed regression model of Task (2) × Block Size (2) × Skill Level (2), with initial distance between hand and object as a covariate for the approach phase data. A similar analysis was performed for the placement phase data, with initial distance between hand and tower or container as the covariate. Table 1 gives the adjusted means and standard errors for all kinematic measures for both phases of both tasks.

Task requirements and kinematic differences. As expected, the degree to which precise movement was required in the task changed the topography of the reach during both the approach and the placement phase. When toddlers approached a block with the intention of later placing it on a tower, peak velocity was reached earlier in the movement, followed by a long deceleration that ended in picking up the block (PPV = 0.42 for tower and 0.50 for throwing). Statistically this was expressed as a main effect of task in the PPV measure, $F(1, 22) = 4.31$, $p < .05$, Cohen's $f = 0.33$. A lower PPV value indicates that the hand spent a longer portion of the reach in slowing its approach to grasp the block. The approach in the tower task also had a shorter movement time than the throwing task, $F(1, 22) = 7.30$, $p < .01$, Cohen's $f = 0.45$ (average MT for tower task was 0.97 s and for throwing was 1.16 s). The finding for movement time was unique in that it is the only measure significant in the opposite

Table 1

Adjusted Means and Standard Errors for All Kinematic Measures in the Precise and the Imprecise Tasks in the Approach Phase and the Placement Phase in Study 1

	Tower (precise)	Container (imprecise)
Approach phase		
Movement time (s)	0.97 (0.05)	1.16 (0.05)**
Straightness ratio	1.81 (0.15)	1.98 (0.15)
Average speed of the reach (mm/s)	319.84 (13.81)	304.50 (13.80)
Amplitude of peak velocity (mm/s)	684.91 (43.77)	615.59 (43.75)
Point of peak velocity	0.42 (0.02)	0.50 (0.02)*
End speed (mm/s)	210.12 (17.66)	245.30 (17.64)
Placement phase		
Movement time (s)	1.80 (0.07)	1.02 (0.07)**
Straightness ratio	1.67 (0.08)	1.72 (0.08)
Average speed of the reach (mm/s)	189.90 (12.18)	347.63 (12.17)**
Amplitude of peak velocity (mm/s)	556.31 (30.53)	706.95 (30.50)**
Point of peak velocity	0.32 (0.02)	0.61 (0.02)**
End speed (mm/s)	79.26 (28.56)	383.85 (28.53)**

*$p < .05$. **$p < .01$.

direction to that predicted. There is no ready explanation for this, and as reported next in the placement phase movement time was significant in the expected direction, with the tower task having a longer placement movement.

Once the block was grasped, the movement to place it on a tower was greatly slowed compared with a throwing movement into the container. Several velocity measures reflected this difference: Tower building had a lower average velocity, $F(1, 27) = 86.24$, $p < .0001$, Cohen's $f = 1.66$; a lower peak velocity, $F(1, 27) = 11.89$, $p < .002$, Cohen's $f = 0.59$; a lower contact or end velocity, $F(1, 27) = 53.32$, $p < .0001$, Cohen's $f = 1.30$; and as in the approach phase, peak velocity was reached earlier in the movement, $F(1, 27) = 83.41$, $p < .0001$, Cohen's $f = 1.63$ for PPV. Duration of movement was also longer for tower building, $F(1, 27) = 74.07$, $p < .001$, Cohen's $f = 1.54$, as would be expected from the slower velocity measures. Figure 2a

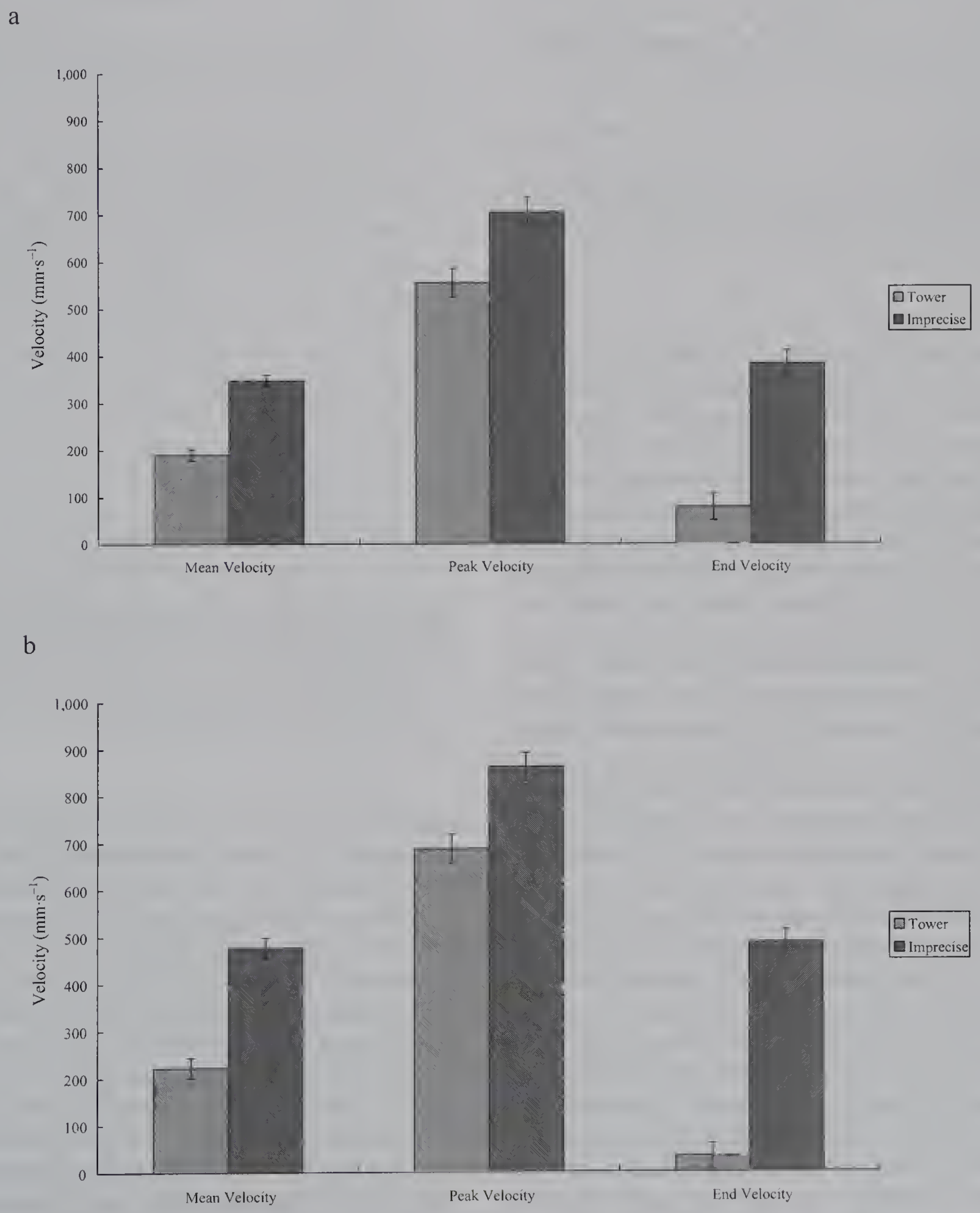


Figure 2. Three velocity measures for the placement phase are displayed for the children at 20 months in Study 1 (a) and a year later in Study 2 (b).
Note. There was consistent slowing of the hand when engaged in a precision task.

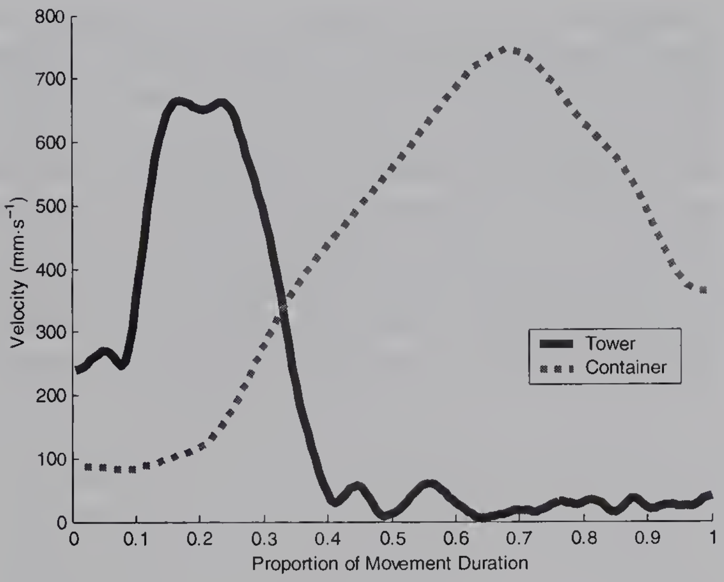


Figure 3. The velocity profile on a single trial from the same child on each task.
Note. This child's hand reached peak velocity earlier when placing a block on the top of the tower compared with throwing a block into the container. Reaches were normalized to the movement time.

displays the group averages for velocity measures, MEANV, PV, and ENDV, showing the consistent slowing of the hand when engaged in a precision task. Figure 3 shows the velocity profile for a single trial from a child on each task, illustrating how the PPV measure varied with the task. This child's hand reached peak velocity earlier, allowing more time for small corrective movements when placing a block on the top of the tower compared with throwing a block into the container.

Skill-level differences in reaching kinematics. Once again, the point of reaching peak velocity in the movement (PPV) proved to be an extremely sensitive measure, $F(1, 27) = 5.28$, $p < .03$, Cohen's $f = 0.37$ (mean PPV was 0.43 for the high-tower group and 0.50 for the low-tower group). The PPV measure is notable because when PV was reached early in the movement, this was followed by a lengthy deceleration as the hand approached the moment where the block was released. A long controlled deceleration reflects good control over the hand movement and would be expected in a skilled performer. Figure 4a shows velocity of a single reach for a high-tower child and a low-tower child when placing a block on a stack. Note the long deceleration limb as the hand approached the point of release for the high-tower child. Although end velocity and average velocity did not differ for high- versus low-tower children, the topography of the reach differed dramatically as shown in Figure 4. We expected to find, but did not, an interaction of skill level with task precision. That is, high-tower builders were expected to move more slowly than low-tower builders when building towers but not when placing blocks into a container, but the PPV difference was still evident when throwing (see Figure 4b). In summary, skill-level differences were reliable only for the placement phase, and the earlier PPV was present in both tasks for the high-tower group.

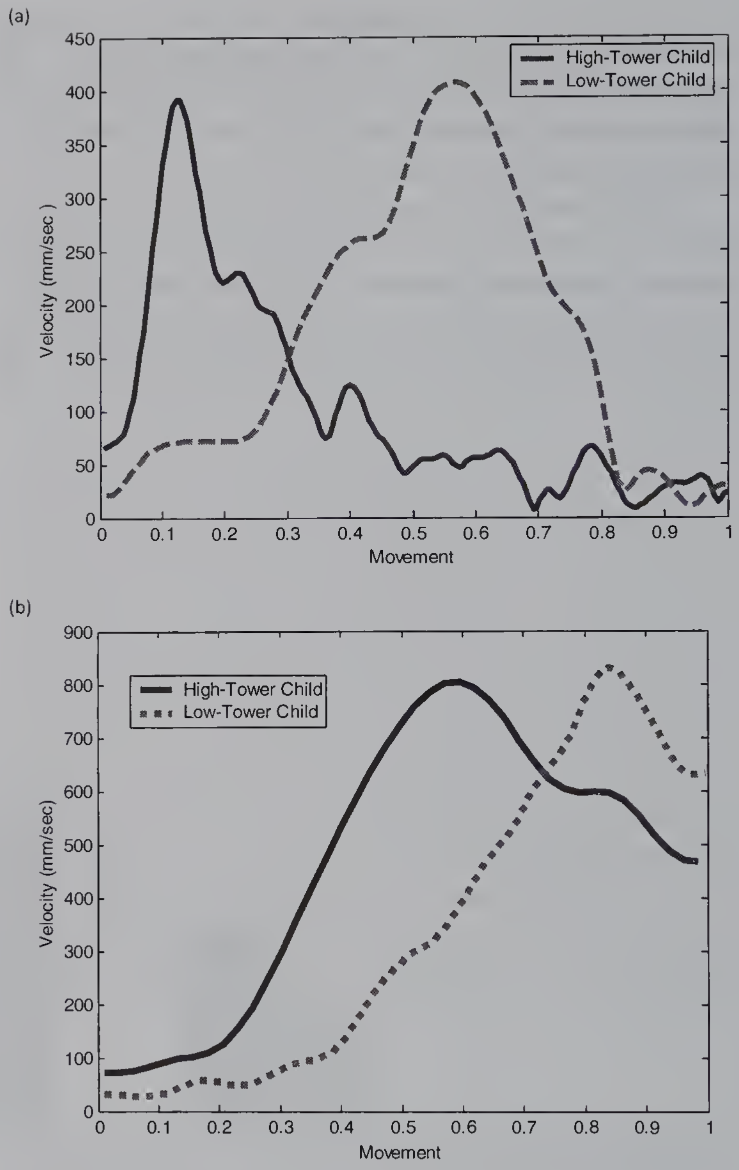


Figure 4. The velocity profile on a single trial of the (a) tower-building task and (b) imprecise task for a high-tower child and a low-tower child.
Note. The hand of the high-tower child reached peak velocity earlier in both cases compared with the low-tower child. Reaches were normalized to the movement time.

The slowed-down approach toward the tower seems to be a necessary component of successfully placing blocks on a tall tower. We hypothesized that the kinematics of the movement toward the tower would change as the stack grew higher. An analysis was carried out on the kinematics of the high-tower group when they were placing blocks low in the tower (under 4 blocks) versus high in the tower (4+ blocks). Contrary to our expectations, children did not change their strategy while building the tower. High-tower children showed greater deceleration when approaching the tower at all levels. Perhaps they represented the functional act of placing a block on the tower as a unit to be repeated, without regard for the changing level of the tower.

Study 2

The purpose of Study 2 was to assess the kinematics of reaching and manipulating blocks a year later in the same tasks that children were given at 18–21 months of age. The issue of greatest interest was whether the skill-level differences apparent at the earlier age would persist a year later when all children could build tall towers. Earlier differences might be transitory if caution in placing blocks on the tower was produced because children whom we designated as high-tower got tested just at the point of learning to master a new skill. In other words, the difference between high- and low-tower children was due to catching some children while they were in the process of achieving new motor control in building block towers. On the other hand, if their kinematics at 18–21 months were indicative of some more general motor skill, the high-tower children would be expected to move with greater skill a year later. That is, if their greater skill at 18 months was a stable characteristic, then group differences would be expected to remain.

Participants

Seventeen of the 31 children who could build at least two-level block towers in Study 1 returned to participate in Study 2. One of the 17 children who was diagnosed as having a developmental delay with recurrent seizure attacks was excluded from the final analysis, leaving 16 children (10 males and 6 females) aged 30–37 months (mean age = 33.98 months, *SD* = 2.37 months). Among the 16 children, 9 could build towers of four or more blocks in their first visit (high-tower group) and 7 could build towers of two or three blocks (low-tower group). All parents who still lived in the area were again contacted by an informational letter describing the study, which was followed by telephone call to see if they were interested in returning to the laboratory. Each child received a book and a certificate as tokens of our appreciation for their participation. All parents or legal guardians of the toddlers signed the informed consents in accordance with the policies of Institutional Review Board of the University of Massachusetts prior to testing.

Results and Discussion

In preliminary analyses, comparisons of the initial distance of the approach and placement phase again revealed a difference between the imprecise and tower tasks. As in Study 1, for the approach phase children's reaches began significantly farther away from the block to be picked up in the imprecise tasks, $F(1, 15) = 25.72$, $p < .0001$. The mean distance was 241 mm (*SE* = 9 mm) for the imprecise task and 188 mm (*SE* = 9 mm) for the tower task. For the placement phase, children's reaches also started significantly farther away from the final release point in the imprecise tasks, $F(1, 15) = 5.00$, $p < .04$. The mean distance was 241 mm (*SE* = 9 mm) for the imprecise task and 212 mm (*SE* = 9 mm) for the tower task. Because the distance the hand traveled could affect kinematic measures, such as MT and PV, we statistically controlled for initial distance as a covariate using mixed regression models to test the differences between tasks, skill level, and block size on each kinematic measure using the SAS program version 8.0 (SAS Institute Inc., Cary, NC).

Regardless of their skill level in Study 1, all children, now between 2.5 and 3.0 years of age, could build tall towers. The mean number of blocks in the stack for the high-tower group was 7.71 and for the low-tower group, 7.11. All reaches used to build towers were averaged and used to represent the tower task. The average number of reaches per child during the approach phase was 9.25 (*SD* = 4.88) for the tower task and 28.88 (*SD* = 9.07) for the imprecise task. The average number of reaches per child during the placement phase was 11.25 (*SD* = 3.89) for the tower task and 29.56 (*SD* = 10.96) for the imprecise task. The data for each kinematic measure were analyzed with a mixed regression model of Task (2) × Block Size (2) × Skill Level (2), with initial distance between hand and object as a covariate for the approach phase data. A similar analysis was performed for the placement phase data, with initial distance between hand and tower or container as the covariate. Table 2 gives the adjusted means and standard errors for all kinematic measures for both phases of both tasks.

Table 2
Adjusted Means and Standard Errors for All Kinematic Measures in the Precise and the Imprecise Tasks in the Approach Phase and the Placement Phase in Study 2

	Tower (precise)	Container (imprecise)
Approach phase		
Movement time (s)	0.77 (0.04)	0.73 (0.04)
Straightness ratio	1.44 (0.06)	1.43 (0.06)
Average speed of the reach (mm/s)	431.30 (17.02)	452.88 (16.83)
Amplitude of peak velocity (mm/s)	787.14 (33.04)	739.70 (32.64)
Point of peak velocity	0.42 (0.02)	0.46 (0.02)
End speed (mm/s)	285.31 (25.83)	346.74 (25.50)
Placement phase		
Movement time (s)	1.59 (0.07)	0.75 (0.07)**
Straightness ratio	1.38 (0.03)	1.36 (0.03)
Average speed of the reach (mm/s)	221.03 (21.39)	476.12 (21.28)**
Amplitude of peak velocity (mm/s)	686.87 (30.61)	861.40 (30.48)**
Point of peak velocity	0.22 (0.02)	0.62 (0.02)**
End speed (mm/s)	32.10 (27.15)	486.77 (26.95)**

**$p < .01$.

Task and skill-level differences a year later. No task differences emerged during the approach phase for these children. Once the block was grasped, however, the movement to place it on a tower was greatly slowed compared to throwing it into the container, similar to findings of Study 1. Several velocity measures reflected this difference: Compared with the throwing task, tower building had a lower average velocity, $F(1, 14) = 138.70$, $p < .0001$, Cohen's $f = 2.93$; a lower peak velocity, $F(1, 14) = 40.98$, $p < .0001$, Cohen's $f = 1.58$; and a lower end velocity, $F(1, 14) = 173.2$, $p < .0001$, Cohen's $f = 3.28$. Peak velocity was reached earlier in the movement, $F(1, 14) = 369.38$, $p < .0001$, Cohen's $f = 4.80$. See Figure 2b for a display of these task differences. Duration of movement was also longer for tower building, $F(1, 14) = 135.84$, $p < .0001$, Cohen's $f = 2.90$, as would be expected from the slower velocity measures.

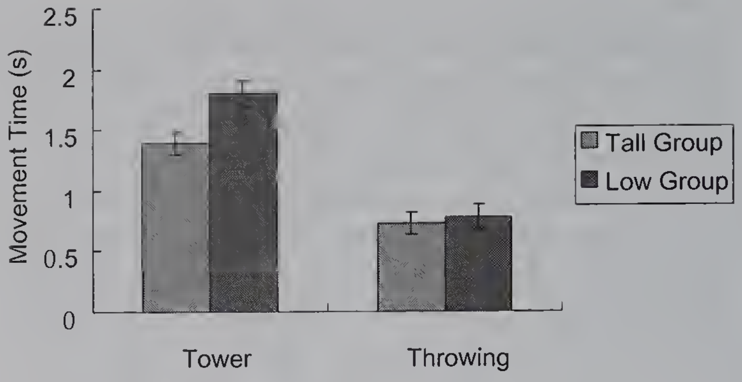


Figure 5. Movement time in the high-tower group and the low-tower group in precise and imprecise tasks in Study 2.

The groups based on skill level, as determined by their tower building during the earlier visit, still had kinematic differences when building towers. Although both groups built equally tall towers in this second visit, the hand movements of the high-tower group were faster. They had a shorter movement time between grasping the block and placing it on the stack but no group difference when throwing blocks into the container (see Figure 5). Statistically this was shown by a Task × Group interaction for MT, $F(1, 14) = 7.44$, $p < .02$, Cohen's $f = 0.63$. In addition, the high-tower group reached peak velocity later in the reach, leaving less time for the hand to decelerate as the block was released from the fingers, $F(1, 14) = 6.91$, $p < .02$, Cohen's $f = 0.61$. This resulted in a higher ENDV for this group, $F(1, 14) = 4.41$, $p < .05$, Cohen's $f = 0.46$. Being able to accomplish a precision task at a higher speed reflects greater mastery in motor control for this group. In Table 3, we list all reliable differences between high- and low-tower groups. The table also lists the PPV for the earlier test session because it is of interest to compare the absolute numbers for this important measure across ages.

Continuity within each child was tested by correlating their kinematic measures at the younger age with those same measures at the older age. All significant correlations were specific to the throwing task. MEANV was significant ($rs = .501$ and $.603$ for approach and placement phases, respectively), and peak velocity was reliable for placement ($r = .53$). Movement time and straightness ratio of the approach phase had significant correlations ($rs = .56$ and $.58$ for MT and SR, respectively). MEANV, MT, and SR during approach were the same three variables that showed consistent improvement over age (see next section and Table 4), and these correlations suggest that the maturing arm/hand control between 2 and 3 years of age has stability within individual children.

Developmental change in reaching kinematics between 20 and 30 months of age. Because the same children were tested about a year apart, developmental changes in reaching kinematics could be compared. Between the 2nd and 3rd years of life manual skill is advancing rapidly, shown by

Table 3
Adjusted Means and Standard Errors for Significant Kinematic Differences Between the High-Tower Group and the Low-Tower Group in the Placement Phase of Studies 1 and 2

	High-tower group		Low-tower group	
	Tower	Container	Tower	Container
Study 1				
Placement phase				
Point of peak velocity*	0.28 (0.03)	0.58 (0.03)	0.36 (0.03)	0.65 (0.03)
Study 2				
Placement phase				
Movement time*	1.39 (0.09)	0.73 (0.10)	1.80 (0.11)	0.78 (0.11)
Point of peak velocity*	0.25 (0.02)	0.66 (0.02)	0.18 (0.03)	0.58 (0.03)
End speed*	52.57 (34.57)	554.62 (35.16)	11.63 (41.45)	418.93 (39.76)

*Significant skill-level difference, $p < .05$.

Table 4
Adjusted Means and Standard Errors for Significant Kinematic Differences Averaged Over Tasks Between Children Tested at 20 Months (Study 1) and Retested at 34 Months (Study 2) in the Approach and Placement Phases

	Younger age (Study 1)	Older age (Study 2)
Approach phase		
Movement time** (s)	1.04 (0.04)	0.70 (0.04)
Straightness ratio**	1.74 (0.06)	1.45 (0.06)
Average speed of the reach** (mm/s)	328.38 (13.93)	416.59 (13.90)
Placement phase		
Movement time* (s)	1.32 (0.06)	1.14 (0.06)
Straightness ratio*	1.63 (0.05)	1.44 (0.05)
Point of peak velocity*	0.49 (0.02)	0.43 (0.02)

Significant age difference: $*p < .05$. $**p < .001$.

children's mastery of fine motor control over fingers and arms to perform precision movements. In order to make comparisons between the first and second sessions, the initial distance the hand traveled during the approach and placement phases had to be taken into account. This distance was shorter when the children were younger, as would be expected because greater arm length and hand size would extend the length of reaches at the older age. For the approach phase, the mean distance traveled was 183.22 mm (*SD* = 38.17 mm) at the younger age and 195.06 mm (*SD* = 55.84 mm) when children were retested. Because the distance the hand traveled could affect kinematic measures, we statistically controlled for initial distance as a covariate using mixed regression models to test the age differences for each kinematic measure using the SAS program version 8.0 (SAS Institute Inc., Cary, NC).

To test for change in movement kinematics between the 2nd and 3rd years of life an ANOVA on Age × Task × Group was performed for approach and placement phases. When children's hands approached a block, the older children had a shorter movement duration, MT: $F(1, 14) = 49.51$, $p < .0001$, Cohen's $f = 1.74$; reaches were straighter, SR: $F(1, 14) = 14.06$, $p < .01$, Cohen's $f = 0.90$; and had a higher average velocity, MEANV: $F(1, 14) = 37.52$, $p < .0001$, Cohen's $f = 1.51$. Once the block was grasped, the older children again took less time to place a block on top of the tower or throw it into a container, MT: $F(1, 14) = 5.58$, $p < .04$, Cohen's $f = 0.54$; they had straighter movements toward the target, SR: $F(1, 14) = 7.86$, $p < .02$, Cohen's $f = 0.65$; and a longer deceleration phase at the end of the reach, PPV: $F(1, 14) = 6.97$, $p < .02$, Cohen's $f = 0.61$. See Table 4 for adjusted means and standard errors for these kinematic measures at both ages. In general, as children approached 3 years of age, their reaches were faster and more efficient in grasping and manipulating the blocks.

General Discussion

Building a tower of blocks is a typical activity of children around 18–24 months of age. It reflects their ability to form a goal and pursue it by repeatedly executing sequential actions. The ability to link perceptual information (in our case shape and size of blocks, height of tower) with the proper motor actions suggests that an overarching goal governed the entire sequence. We examined the kinematics of the approach phase to grasp a block and then the movement to place the block on the tower. Both phases were compared to a more casual task of picking up blocks to put them into a large open container. The approach phase kinematics were expected to differ if children were thinking ahead to their goal after they grasped the block, whereas the kinematics of placement were assumed to reflect the care with which children positioned the block for release (Johnson-Frey, McCarty, & Keen, 2004). This two-stage task of first grasping then manipulating the blocks offers a means of assessing how far ahead intentional actions are planned. Johnson-Frey et al. (2004) had adults reach for a block and subsequently lift and hold it, or transport it to a new location. When adults were anticipating a transport movement, the initial reach-to-grasp duration was affected. The authors concluded that the approach phase was not planned in isolation but was represented in the complete action sequence, including future movements once the block was in hand. Our data show that toddlers' reach-to-grasp actions also indicate planning of the entire sequence rather than executing movements in isolation.

For the approach phase, we found that toddlers reached for blocks differently depending on the task they intended to subsequently perform. In adults, this difference shows up for many kinematic measures, such as total movement time, average speed, peak speed, and when peak speed is reached (Johnson-Frey et al., 2004; Marteniuk et al., 1987), but only the timing of peak speed and the movement time were reliable for toddlers. It was expected that average and peak speed would be lower for the precision task as this was found for infants (Claxton et al., 2003) as well as for adults. Adults and infants had highly similar tasks: Pick up an object and either fit it into a tight hole or throw it into a container. Toddlers may not have viewed the tower-building versus "clean-up" tasks as differing very much in precision compared with the previous "fit it or throw it" tasks. Future studies with toddlers should test them on tasks more similar to what adults did. Nevertheless, as predicted from the adult data, peak speed was reached earlier in the approach for tower building. This task difference suggests that toddlers are beginning to plan sequential stages of the action from the very beginning of movement. Although there is no obvious need for the hand to slow in its approach to pick up the block, the intention to subsequently perform a precision task appears to affect hand speed in the first stage of this two-stage action. We conclude that for this age PPV is a more sensitive measure of the added care used when approaching blocks to be used in building a tower.

In the placement phase, toddlers exhibited numerous kinematic differences between the two tasks. When building a tower, all three measures of velocity (peak, average, and contact velocity) reflected slowing of the hand throughout the movement as blocks were lifted and added to the stack, compared with being lifted and released into an open container. Thus, children appeared to hold their goal in mind when performing sequential actions involving several items, and exhibited lower velocity when engaged in a precision task just as adults do (Berthier, Clifton, Gullapalli, McCall, & Robin, 1996; Bryden & Roy, 1999; Johnson-Frey et al., 2004; Wu et al., 2005).

The timing of peak velocity proved to be more sensitive to task demands and to skill-level differences than other kinematic measures for this age group. Why should this be so? Presumably the point at which peak speed is reached, or PPV, reflects how control demands are exercised during the movement. The beginning of a movement is primarily devoted to transporting the limb whereas the later parts are controlling the final approach and securing an appropriate grasp in the approach phase and release in the placement phase. The later parts are more dependent on visual feedback for adjusting the details of the movement. In infants' reaching, movements are produced in a series of submovements (movement units), each with a distinct acceleration and deceleration phase (von Hofsten, 1979, 1991). As infants come to master reaching, the largest unit will be situated at the beginning of the movement with one or several smaller units following toward the end. The large unit at the beginning is devoted to the transport whereas the smaller units are devoted to the grasp or release. At the end of the movement, the object and the reaching hand are simultaneously visible and control can therefore be quite precise. If precision demands are high, peak speed will occur earlier to allow time for precise control. On the other hand, if precision demands are low, peak speed will occur later because the movement before peak speed is reached is focused on transport with less need for slowed down movement.

As noted earlier, peak speed was reached earlier in the precise task and the hand noticeably slowed for the final approach. Although all children displayed this pattern, it was amplified in children who were more skilled at tower building. The high-tower group reached peak speed earlier than the less skilled group when they were initially learning to build a tower. Although their approach to pick up the blocks did not differ, the high-tower children had a more exaggerated deceleration as they placed blocks on the stack. We interpreted this to mean that these children employed the same strategy that adults do when engaged in a precision task, namely, reach peak speed early in the movement in order to slow down before reaching the point where greatest control is needed. For adults, this pattern is shown when picking up a very small object (Berthier et al., 1996), or placing an object into a tight-fitting space (Marteniuk et al., 1987). In this study this same pattern appeared in highly skilled children, suggesting that they realized that placing a block on a tower required care in keeping the entire stack balanced.

Unexpectedly, toddlers who were more skilled also showed this pattern of reaching peak velocity early when releasing blocks into the container when there was no need for extra care. Children of this age who can construct high block towers have probably achieved this skill recently, according to the Bayley norms. We propose that in performing this new skill, the care and attention it required spilled over into the imprecise task that involved the very same blocks (Boudreau & Bushnell, 2000). We suggest that children who have recently acquired the ability to construct tall towers adopted similar movement characteristics for manipulating the blocks in any task. The fact that the same blocks were used in both tasks may have facilitated this tendency. As children become older, building block towers is less challenging and their manipulation kinematics would be expected to differ with task difficulty. When retested a year later, the high-tower group still had kinematic differences from the low-tower group when building a tower but not when placing blocks into an open container.

As toddlers, the high-tower children moved more slowly than low-tower children as they placed blocks on a tower, but a year later they moved more quickly. This change appears to reflect an evolution in skill of tower building. When first learning to successfully balance several blocks in a tower, going slowly is associated with better success. Once mastery is achieved, these children show their superior motor control by going faster, just as highly skilled drivers or skiers can go faster than those with less skill. The fact that throwing velocity did not differ for the groups is evidence that the high-tower children do not move faster in general, but only in the precision task. It would be interesting to have tested other precision tasks because we do not know if their skill was specific to tower-building or was characteristic of their approach to precision tasks in general.

We expected that greater maturity in manual skills would be evident in many kinematic measures when a subset of children from the original study was brought back a year later. The same pattern of task differences prevailed at both ages: Average velocity, peak velocity, and velocity at contact were lower when building a tower than when throwing blocks into a container, and peak velocity was reached earlier in the movement. Within this similar velocity profile, age differences emerged. Reaches at the older age had a faster average velocity as the hand approached the object, and movements had straighter trajectories and shorter durations for both approach and placement phases. Moving faster and straighter toward one's goal in both phases of movement is exactly the expected result of greater motor maturity and would be predicted from others' findings (Contreras-Vidal, 2006; Olivier & Bard, 2000). Konczak and Dichgans (1997) found similar changes in a longitudinal sample of children tested between 5 months and 3 years of age. The task in Konczak and Dichgans was a simple reach for an object held within the child's reach at his or her shoulder height; there was no placement phase in this study. They reported progressive straightening of the path, and increasing peak velocity, with reaches approaching the stereotypic adult pattern of arm kinematics by the 3rd year of life. Of special interest is their finding that the timing of peak velocity moves earlier in the reach, just as in our data, with greater maturity. Task difficulty was not manipulated in this study, but Carrico and Berthier (2008) did test 15-month-olds in the approach phase of reaching for big and small objects. The small object required a pincer grip. Their results mirror ours in several ways. The approach for pickup of the smaller object (a Cheerio) had slower average speed, longer movement time, and peak velocity was reached earlier in the movement, compared with the approach for the larger object (rubber ball).

The change in kinematic measures over age allows us to distinguish two factors included in the development of a skill. The first is about mastering precision requirements, an important aspect of which is allocating relatively more time to visual control of the final part of the movement. When the children in the low-tower group had mastered building a high block tower a year later, they showed very early PPV, the signature measure for extended time at the end of the reach. A second factor included in skill development is making the action more efficient. This includes shorter movement duration, straighter trajectories, and higher average velocity which all children showed when retested a year later. Most striking in the high-tower children a year later, PPV moved later in the reach so that the hand was moving faster when placing blocks on the stack. This higher velocity when performing a precision task reflects greater efficiency for this group.

In summary, kinematic measures of toddlers' performance yielded insight into how they approach and execute tasks varying in difficulty. In particular, the timing of peak speed in relation to the endpoint of the movement was sensitive to children's level of skill and to task demands. Reaching peak velocity earlier in the movement in demanding tasks gives rise to a longer period of deceleration and thus better visual control before the hand arrives at the moment of either contacting the block in the approach phase or releasing it in the placement phase. This pattern was evident at both ages when children reached for a block when the upcoming action required precision and balance. Not surprisingly, this same kinematic pattern was demonstrated during the actual placement of the block on the tower, compared with transferring it to a container. Finally, the sensitivity of this movement timing extended to differentiating among children of different skill levels, a difference that remained stable over 1 year between test sessions. By using two-stage tasks that involved sequential movements, we demonstrated that toddlers engaged in movement planning beyond the available perceptual information and incorporated the final goal of the entire sequence into that planning.

References

Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development*. New York: Psychological Corporation.

Berthier, N. E., Clifton, R. K., Gullapalli, V., McCall, D. D., & Robin, D. J. (1996). Visual information and object size in the control of reaching. *Journal of Motor Behavior, 28*, 187–197.

Boudreau, P., & Bushnell, E. (2000). Spilling thoughts: Configuring attention trade-off in infants' goal directed actions. *Infant Behavior & Development, 23*, 543–566.

Bryden, P. J., & Roy, E. A. (1999). Spatial task demands affect the extent of manual asymmetries. *Laterality, 4*, 27–37.

Carrico, R. L., & Berthier, N. E. (2008). Vision and precision reaching in 15-month-old infants. *Infant Behavior & Development, 31*, 62–70.

Claxton, L. J., Keen, R., & McCarty, M. E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science, 14*, 354–356.

Contreras-Vidal, J. L. (2006). Development of forward models for hand localization and movement control in 6- to 10-year-old children. *Human Movement Science, 23*, 634–645.

Gentiner, D. R. (1983). Keystroke timing in transcription typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 95–120). New York: Springer-Verlag.

Gesell, A. H., & Amatruda, C. S. (1954). *Developmental diagnosis: Normal and abnormal child development*. New York: Harper.

Johnson-Frey, S., McCarty, M., & Keen, R. (2004). Reaching beyond spatial perception: Effects of intended future actions on visually guided prehension. *Visual Cognition, 11*, 371–399.

Konczak, J., & Dichgans, J. (1997). The development toward sterotypic arm kinematics during reaching in the first 3 years of life. *Experimental Brain Research, 117*, 346–354.

Larochelle, S. (1983). A comparison of skilled and novice performance in discontinuous typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 67–94). New York: Springer-Verlag.

Marteniuk, R. G., MacKenzie, C. L., Jeannerod, M., Athenes, S., & Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology, 41*, 365–378.

Olivier, I., & Bard, C. (2000). The effects of spatial movement components precues on the execution of rapid aiming in children aged 7, 9, and 11. *Journal of Experimental Child Psychology, 77*, 155–168.

Temprado, J., Della-Grasta, M., Farrell, M., & Laurent, F. M. (1997). A novice-expert comparison of intra-limb coordination subserving the volleyball serve. *Human Movement Science, 16*, 653–676.

von Hofsten, C. (1979). Development of visually guided reaching: The approach phase. *Journal of Human Movement Studies, 5*, 160–178.

von Hofsten, C. (1991). Structuring of early reaching movements: A longitudinal study. *Journal of Motor Behavior, 23*, 280–292.

Weir, P. L., MacDonald, J. R., Mallat, B. J., Leavitt, J. L., & Roy, E. A. (1998). Age-related differences in prehension: The influence of task goals. *Journal of Motor Behavior, 30*, 79–89.

Wu, C.-Y., Lin, K.-C., Lin, K.-H., Chang, C.-W., & Chen, C.-L. (2005). Effects of task constraints on reaching kinematics by healthy adults. *Perceptual and Motor Skills, 100*, 983–994.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1859–1875

Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children

Géraldine Legendre
Johns Hopkins University

Isabelle Barrière
Brooklyn College, CUNY,
& Yeled v'Yalda Research Institute

Louise Goyet and Thierry Nazzi
Université Paris Descartes

Two comprehension experiments were conducted to investigate whether young French-learning children ($N = 76$) are able to use a single number cue in subject–verb agreement contexts and match a visually dynamic scene with a corresponding verbal stimulus. Results from both preferential looking and pointing demonstrated significant comprehension in 30-month-olds with no preference for either singular or plural. These results challenge previous claims made on the basis of English and Spanish that comprehension of subject–verb agreement expressed as a bound morpheme is late, around 5 years of age (V. A. Johnson, J. G. de Villiers, & H. N. Seymour, 2005; A.-T. Pérez-Leroux, 2005). Properties of the adult input were also analyzed. Possible implications for theories of syntactic acquisition are discussed.

Many languages encode a distinction between singular and plural reference, and do so in a variety of ways. This distinction is sometimes encoded redundantly, resulting in grammatical agreement. In English, for example, the number of a referent noun is typically encoded in a determiner entering a dependency with presence versus absence of an –*s* morpheme on the noun: *A boy* (singular) versus *Some boys* (plural). When such nouns occur in subject position they enter a dependency with a verb and the number of the referent noun is also reflected on the verb; however, presence of an –*s* morpheme on the verb marks singular while its absence marks plural: *A boy jumps* versus *Some boys jump*. The relation between the –*s* morpheme and the number category across agreeing constituents is therefore, not a simple one. It is further complicated by the fact that the –*s* morpheme has three phonological realizations, depending on properties of the preceding segment: /əz/ after voiced and voiceless consonants corresponding to segments like ''s,'' ''z,'' ''ch,'' and ''j'' (e.g., *buzzes*), /z/ after vowels and remaining voiced consonants (e.g., *goes*, *jogs*), and /s/ after remaining voiceless consonants (e.g., *jumps*). Despite its phonological variation, –*s* is highly regular in the language because it applies to all verbs, apart from modals (e.g., *can*, *may*, etc.) and auxiliaries which involve additional phonological changes to the roots/stems (e.g., singular *has*, *does*, *is* vs. plural *have*, *do*, and *are*).

The task for a child acquiring any language with an agreement system is at least twofold: She must acquire the expression of number distinctions on nouns and verbs and the principle of agreement across syntactic categories and functions (e.g., subject) that holds in her language. Prerequisites appear to be in place by age 2. For example, the prerequisite category of ''noun'' is supported by studies based on productivity of use of appropriate morphemes (Tomasello & Olguin, 1993) and that of

---

The research described here was supported by Grant BCS0446954 from the National Science Foundation. We are grateful to all of the children who participated in these studies and to their parents. We thank Davis Anderson, Elika Bergelson, Lauren Cash, Joyce Chen, Henri Coulaud, Jenny Culbertson, Cristina Escalante, Malki Grusman, Fay Halberstam, George Hollich, Sarah Kresh, Monica Lopez-Gonzalez, Ariana Martohardjono, Tabitha Moses, Sylvain Mottet, Oni Okolo, Ludmila Pukhovich, Lilia Rissman, Susan Smedesdran, Reiri Sono, Sara Tincher, Connie Yau, and Erin Zaroukian for assistance in the preparation of the stimuli, data collection, and/or coding, and Hilary Liberty for help on statistical analyses.

Correspondence concerning this article should be addressed to Géraldine Legendre, Department of Cognitive Science, Krieger Hall, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218. Electronic mail may be sent to legendre@jhu.edu.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0018

''verb'' by studies on novel verb learning (Bernal, Lidz, Millotte, & Christophe, 2007). Asymmetrical omission in production indicates that 2-year-olds already distinguish a category of ''subject'' (Valian & Aubry, 2005). Even 17-month-olds show sensitivity to word order suggestive of a subject category in a preferential looking experiment, correctly interpreting the subject in ''Where's Big Bird tickling Cookie Monster?'' by picking the video of Big Bird tickling Cookie Monster, not Cookie Monster tickling Big Bird (Hirsh-Pasek & Golinkoff, 1996).

Several studies have explored children's comprehension and production of subject–verb agreement. First, studies of language production fail to find evidence that 2-year-olds have knowledge of subject–verb agreement. The first multiword utterances spontaneously produced by young children acquiring English do not exhibit its systematic use. Mastery of third-person singular agreement (defined as 90% of use in obligatory contexts) takes place between the ages of 26 and 46 months (Brown, 1973; De Villiers & de Villiers, 1973). In a sentence completion task with nonce verbs requiring the suffix /əz/ (e.g., *nazzes*), even 7-year-olds were shown to have difficulty producing this agreement, succeeding only about 48% of the time (Berko, 1958). In another study comparing the three phonological instantiations of the –*s* morpheme in English, /z/ was found to be the first one to be productively used for marking plural in nouns in an elicitation task, at the age of 6 (Anisfeld & Tucker, 1967). Overall, these production studies show poor performance levels that, moreover, appear to be sensitive to the task used (elicited vs. spontaneous production).

Second, subject–verb agreement has been evaluated using tests of language comprehension, which only found evidence of emergence late in acquisition (Johnson et al., 2005). In a picture selection task using auditory stimuli designed to phonologically mask the number on the noun by using verbs beginning in *s*– such that the child would have to rely on the verbal marking to provide a correct answer, they found that 5- and 6-year-olds, but not 3- and 4-year-olds, were able to select appropriate pictures when hearing *The duck swims in the pond* or *The ducks swim in the pond*. In every age group tested, accuracy was higher for singular –*s* than for plural (null morpheme). The study was replicated in (Caribbean) Spanish, with similar results (Pérez-Leroux, 2005). In that language the pronominal subject is characteristically omitted with the consequence that plural number of the subject may be solely encoded as a suffixal –*n* on the verb (e.g., singular *Nada en el charco* ''[It] swims in the pond'' vs. plural *Nadan en el charco* ''[They] swim in the pond''). The pattern is completely regular in the language and is the same in all varieties of Spanish. Three- to 4½-year-olds performed at chance level in both singular and plural conditions. Older children (4;8–6; 6-year-olds) performed at chance level in the singular condition (null morpheme) but selected the matching picture 67% of the time in the plural condition (–*n*). While the results for Spanish- and English-learning children showed a similar pattern of no comprehension under the age of 4 and similar accuracy level (60%–70%) for children aged 5 and 6, their accuracy level with regard to number was the opposite in the two languages. This was interpreted by Pérez-Leroux (2005) as evidence that children were in fact sensitive to *overt* marking (–*s* on third singular verbs in English, –*n* on third plural verbs in Spanish).

The conclusion of a late emergence of subject–verb comprehension based on Pérez-Leroux's (2005) results is, however, challenged by an independent claim that Spanish-learning children show a preference for singular marking (Childers, Fernandez, Echols, & Tomasello, 2001). In a pointing task using colored drawings of a girl or two children performing actions of walking, singing, sleeping, running, etc., 36-month-old (but not 30-month-old) Chilean Spanish-learning children showed a significant preference for third-person singular auxiliary forms (e.g., *Está corriendo* ''is running'') over plural forms (e.g., *Están corriendo* ''are running''). A trend toward significance in the same direction was found with lexical verbs (e.g., singular *Come los sandwiches* ''eats the sandwiches'' vs. plural *Comen* ''eat''). A study of child-directed speech in Guatemalan Spanish by a subset of the authors is mentioned as evidence that there is a disproportionate number of third-person singular forms in Spanish input to young children, which may explain the bias for singular forms in the experimental results.

Another challenge to the claim of late comprehension of subject–verb agreement comes from a study originally designed to explore comprehension of number marking rather than agreement per se, but suggesting that young children are particularly sensitive to verbal marking (Kouider, Halberda, Wood, & Carey, 2006). In that preferential looking study, 24-month-olds (but not 20-month-olds) looked longer to the matching screen when they heard audio stimuli with several markers of number which included verbal marking (e.g., *There* **are** *some blickets, There* **is a** *blicket*). When only one nominal cue was provided (e.g., *Look at the blickets*, *Look at the blicket*) only 36-month-olds succeeded at the matching task. Detailed looking-time analyses suggested that the most important cue was the distinctive auxiliary (*are* vs. *is*). These results were replicated with familiar nouns using a manual search paradigm and interpreted as evidence that lexical distinctions (*is* vs. *are*) are easier to acquire than bound morphology (*–s*) because the auxiliaries are more salient than a single bound morpheme (Wood, Kouider, & Carey, 2009). They are also more frequent than *–s* marking on a verb in English. Indeed, relative frequencies in the adult input based on a study of two corpora (Nina: 2,835 utterances, Suppes, 1974; Naomi: 705 utterances, Sachs, 1983) reveal that *is* occurs in 26% of adult utterances versus 7% for *–s* (Zaroukian, 2009).

Third, subject–verb agreement was also evaluated through preference studies in younger children given that absence of comprehension does not preclude awareness of agreement. Nineteen-month-olds acquiring English revealed a sensitivity to grammatical subject–verb agreement in *–s* in a headturn preference study, based on grammatical stimuli like *The boy bakes bread* and *The boys bake bread* versus ungrammatical **The boy bake bread* (Soderstrom, 2002; Soderstrom, Wexler, & Jusczyk, 2002). However, these young children were unable to distinguish grammatical from ungrammatical double marking on noun and verb (**The boys bakes bread*) and double verbal marking (**The boy does bakes bread*). This is indicative of some sensitivity to the grammatically appropriate number of *–s* morphemes as well as their placement, hence to the English grammar, rather than a mere sensitivity to and perceptual preference for inflected verbs over uninflected ones (Soderstrom, 2008).

Seeking to tease apart sensitivity to marking from sensitivity to word order as well as the role of lexical knowledge and function words, Soderstrom, While, Conwell, and Morgan (2007) tested young children' ability to track *–s* marking (e.g., verbal as in *The baker bakes bread* or nominal as in *The bakers bake bread*) in passages involving six highly familiar nouns and verbs while manipulating the position of these content words. They found that 16-month-olds were sensitive to grammatically appropriate placement of *–s* on highly familiar nouns and verbs when adjacent to function words (e.g., auxiliaries, prepositions, and determiners). Their study confirms the crucial role played by function words in children' initial formation of grammatical knowledge (Höhle & Weissenborn, 2003; Shady, 1996; Shi, Marquis, & Gauthier, 2006; Shi, Werker, & Cutler, 2006) and provides evidence of the facilitating role played by known content words.

Overall, these last studies suggest that while an early sensitivity to the presence of morphemes encoding number distinctions as well as agreement can be detected relatively early—at least in English—comprehension/production of agreement encoded as a bound morpheme takes a long time to acquire. At present, the reasons for such a lag are unclear. One factor may be the methodology used in the comprehension/production studies, an issue we will discuss in more detail later on. Other factors may be grounded in the properties of the languages tested so far and/or the input received by the children. For example, the fact that one number category is null in these languages (i.e., plural in English, singular in Spanish) may have caused children to focus their attention onto the overt one. Acquisition may also have been influenced by adults' disproportionate use of singular verbal forms around young children.

There is another reason why comprehension might be delayed in English and Spanish. In both languages, overt verbal agreement is marked by a single consonant in (word-final) coda position. Numerous studies on word recognition have found evidence that perception of consonants is affected by their position in a syllable and in a word, both in adults (e.g., Redford & Diehl, 1999) and infants (e.g., Swingley, 2003; Zamuner, 2006; though see Nazzi & Bertoncini, 2009; Swingley, 2009, for positive evidence of coda consonant processing). Swingley (2009) characterizes some of the challenge as follows: ''Word-final consonants are, in general, less clearly articulated; they are heard only after perception of the initial parts of the word has led children to consider an interpretation; and they enjoy less of the benefit of membership in dense phonological neighborhoods.'' Thus, if agreement marking occurred in (word-initial) syllable onset position, earlier comprehension may be found.

French has a subparadigm of agreement of this very sort. While 10% of French verbs have retained phonological encoding of person–number distinctions on the verb itself, the remaining 90% (known as conjugation class I, ending in *–er* in the infinitive) have lost it despite its retention in the spelling. This results in near-complete homophony of the verb form, e.g., /dãs/ ''I, you, he, she, they dance'' (with the exception of first-person plural ''we'' form in *–ons* /dãsõ/ and second-person plural ''you'' form in *–ez* /dãse/, which are extremely rare in child-directed speech). In connected speech, these homophonous verbal forms can be distinguished via subject pronouns under certain conditions. Whenever the verb is consonant-initial the pronunciation of the third-person singular pronoun *il* ''he'' versus plural pronoun *ils* ''they'' is identical, for example, *Il(s) danse(nt)* /i(l)dãs/ ''he/they dance'' though speakers idiosyncratically vary in realizing /l/, independently of number. /i(l)/ thus behaves as a default form in French.

However, for vowel-initial verbs in this conjugation class (e.g., *embrasser* ''to kiss''), which represent about one fourth of the verbs of this class (702 out of 2,673 verbs or 26.3% according to the Manulex ''child'' database, Lété, Sprenger-Charolles, & Colé, 2004; 1,062 out of 4,655 verbs or 22.8% according to the Lexique ''adult'' database, New, Pallier, Ferrand, & Matos, 2001), the phonological fusion of the subject pronoun with the verb results in distinct phonological forms in the third-person singular versus plural. For example, singular *Il embrasse* ''he kisses'' is pronounced /ilãbras/. Its plural counterpart, *Ils embrassent* ''they kiss,'' has two minimally different pronunciations of the pronoun that idiosyncratically vary across speakers: /i**z**ãbras/ and /il**z**ãbras/. In either case, plural marking involves a *liaison* ''linking'' or latent consonant surfacing only with vowel-initial verbs in the immediate presence of a subject pronoun, plus resyllabification whereby the plural marker /z/ finds itself in *onset position of the second syllable*, e.g., /i(**l**).**z**ã.bras/ (Encrevé, 1988; Tranel, 1996). This and other special properties have led a growing number of linguists to propose that French subject pronouns are in fact best analyzed as mere agreement markers (Auger, 1994; Culbertson & Legendre, 2008; Fonseca-Greber & Waugh, 2003; Jakubowicz & Rigaut, 1997; Lambrecht, 1981; Miller, 1992; Pierce, 1992).

The present study, therefore, tests the hypothesis that this French subject–verb agreement subsystem will be comprehended earlier than the English and Spanish type, possibly as early as 24 months of age. In Experiment 1, we built on growing evidence that the intermodal preferential looking paradigm (IPLP; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1996) is an established method for investigating early comprehension of syntax (e.g., Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008; Fisher, 2002; Lidz, Waxman, & Freedman, 2003; Naigles, 1990). In Experiment 2, we exploited recent successful attempts at incorporating pointing as a way of engaging young children (Bernal et al., 2007; Brandone, Pence, Golinkoff, & Hirsh-Pasek, 2007; Dittmar et al., 2008; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). In both experiments, we followed Soderstrom et al. (2007) by only presenting our subjects with verbs known to them, based on parental report. We expected familiarity to facilitate comprehension.

Specifically, we asked whether young French-learning children from two age groups (24 and 30 months) are able to map an auditory stimulus consisting of singular or plural combinations of a subject pronoun + familiar transitive verb + nonce object noun (e.g. *Il embrasse le taque* ''he kisses the tak'') with visually displayed dynamic scenes involving one or two actors.

The present study also investigates whether frequency might play an additional role, as suggested in Childers et al. (2001). While the target type of subject–verb agreement is found only when the subject pronoun *immediately* precedes the verb its actual manifestation in the adult input is unknown. Large samples of French adult input were analyzed in Experiment 3 with the goal of formulating possible implications of our experimental results for current theories of syntactic acquisition.

Experiment 1

In Experiment 1, we tested comprehension of subject–verb agreement by investigating the ability of 24- and 30-month-old children to match an orally presented utterance to the appropriate video among a pair of actions presented simultaneously. The speech stimuli instantiated a common pattern of subject–verb number agreement in French, that is, a lack of phonological singular–plural distinction on the verb itself but a marking through liaison between the subject pronoun and the vowel-initial verb, resulting in number agreement being expressed as a single consonant in syllable onset position. The task used was the IPLP because it places relatively few demands on children by measuring comprehension through looking time, thus without requiring them to either make meta-linguistic judgments or even perform an action such as pointing, which requires more complex decisional and motor planning.

Method

Participants

Participants included 20 monolingual French-speaking 24-month-olds (10 males, 10 females; M = 23 months, 28 days; range = 22 months, 18 days to 26 months, 3 days), and 20 monolingual French-speaking 30-month-olds (10 males, 10 females;

M = 30 months, 27 days; range = 28 months, 19 days to 35 months, 14 days). An additional seven 24-month-olds were tested but excluded from the experiment, due to fewer than four analyzable trials (1), excessive side bias (> 85% looking to one side; 4), not contributing data to all conditions (1), and video file problem (1). An additional seven 30-month-olds were tested but excluded from the experiment, due to fussiness (2), excessive side bias (2), or for not contributing data to all conditions (3). Children were all tested in Paris. They were recruited from a database of parents who volunteered to participate in child development studies and came from diverse socioeconomic backgrounds.

Materials

Video stimuli Two 8-year-old boys recruited for this purpose were filmed performing 1 of 10 simple actions on unfamiliar objects for which the French children tested did not have a name. For each action, either one boy performed the action alone while the other boy was standing immobile next to the first boy (singular video), or the two boys performed the action together and simultaneously (plural video). Figure 1 shows characteristic still images from the actual videos used for the kissing action. For each action, different unfamiliar objects were used in the singular and plural conditions (a total of 20 unfamiliar objects were used). Thus, the same action was performed on different objects by the single boy (singular video) versus the two boys (plural video). All video sequences lasted 6 s.

Audio stimuli. All speech stimuli were recorded by the same female speaker using moderate child-directed speech. Ten verbs, referring to the ten actions of the videos, were used: *accrocher* "to hook," *allumer* "to switch on," *apporter* "to bring," *arrêter* "to stop," *attacher* "to tie," *attraper* "to catch," *embrasser* "to kiss," *enlever* "to remove," *essuyer* "to wipe," and *ouvrir* "to open." These verbs were chosen because they fulfill several important constraints. They are transitive and belong to the dominant conjugation class (Class I), except for *ouvrir* "to open," which nevertheless follows the same phonological pattern; third-person singular and plural forms of the verb are homophonous. They have been found to be known by many children in previously collected MacArthur Communicative Development Inventory (MCDI) data (Kern, 2003; Nazzi, 2005; Nazzi, Floccia, Moquet, & Butler, 2009; Nazzi & Pilardeau, 2007) and in a large cross-sectional corpus (Le Normand, 1986). All verbs are vowel initial, with phonologically identical third-person singular and third-person plural forms, number agreement being only signaled by liaison between the pronoun and the verb. Of the two pronunciations of the plural pronoun (/iz/, /ilz/), the native speaker who recorded the stimuli consistently used the former.

The verbs were embedded in short sentences consisting of the third-person subject pronoun in either singular (*il* "he") or plural (*ils* "they") form, the verb, and an NP (determiner *le* "the" + noun). Pseudo-words rather than real nouns were used as direct objects of the transitive verbs to refer to the unfamiliar objects. We used 20 different pseudo-words (one for each of the unfamiliar objects), which all had a consonant–vowel–consonant structure (e.g., /gef/). The decision to use unfamiliar objects and pseudo-words was made on several grounds, including neutralizing possible lexical noun knowledge effects across subjects, eliminating possible effects of the predictability of the object reported in imitation tasks (Valian, Prasada, & Scarpa, 2006), as well as a possible but problematic distributive interpretation associated with singular phrases/sentences. A sentence like *Il embrasse la poupée* "He is kissing the doll" matches both a scene in which only one boy kisses the doll and a scene in which two separate boys are kissing the doll (see Kouider et al., 2006, for a similar argument). In contrast, the use of distinct pseudo-nouns as objects (e.g., *embrasser le voube* "to kiss the voub" vs.

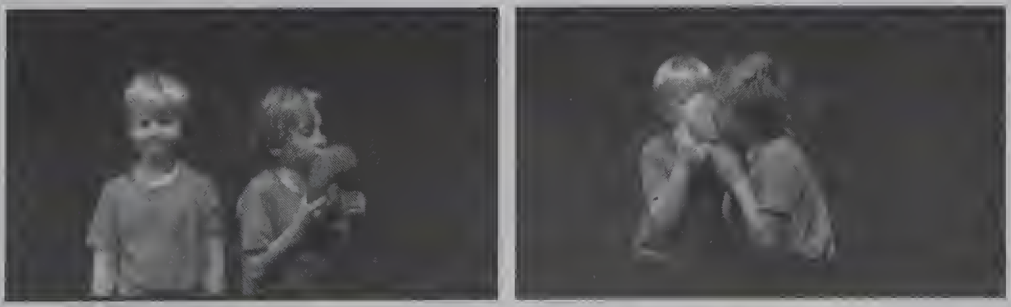


Figure 1. Video stimuli in the singular (left screen) and plural (right screen) conditions.

embrasser le taque "to kiss the tak") refer to events that are distinguished on the basis of the unfamiliar objects involved. A proficient user of French would need to make use of their knowledge of *il* (singular) versus *ils* (plural) to assign a singular or plural interpretation to *il embrasse le voube* and *ils embrassent le taque*.

Procedure

Prior to coming to the lab, parents were asked to name which of the ten verbs used in the experiment their child knew. They were not asked whether their child knew specific forms of the verbs. Right before their visit, parents also completed a French version of the MCDI questionnaire (Fenson et al., 1993; Kern, 2003) modified by the authors to include more verbs and to break down pronominal forms into subcategories to allow for correlation analyses (see below). The MCDI is a parent-friendly, successful, and reliable tool for measuring lexical and function words a young child understands and/or produces spontaneously. MCDI questionnaires were returned for 18 (of 20) 24-month-olds and 17 (of 20) 30-month-olds.

Each child was tested individually in a soundproof booth for about 10 min using the IPLP. Children sat on a parent's lap during the whole session. The parent, wearing dark glasses rendered completely opaque by covering the inside of the lenses with thick paper, was blind to the video stimuli. Each child was presented with 6–8 trials (depending on the number of test verbs reported to be known by their parent). All trials consisted of the sequence of events described in the following paragraph.

The child's attention was first centered by flashing a central light located between the two presentation screens (the child facing the whole display). Once the child looked at the central light, the baseline phase started, in which two videos were simultaneously displayed on two screens in silence for 6 s. Both videos depicted the same action, one being the singular video and the other being the plural video. After this video presentation, the central light was flashed again. Once the child looked at it, the test phase started, consisting of the audio presentation of a speech stimulus matching one of the two scenes, immediately followed by a 6-s-long presentation of the same two videos presented during the baseline phase. Once the test phase was over, a 3-s eye-catching video (a "dance" performed by the two boys) was displayed on the screen of the matching video to mark the end of the trial (following Kouider et al., 2006). Previous studies have shown that children tend to look longer at the scene matching the speech stimulus during test compared to baseline (e.g., Golinkoff et al., 1987; Naigles, 1990).

For half of the trials, the speech stimulus corresponded to the singular video, while it corresponded to the plural video for the other half of the trials. Moreover, the side of the matching video was counterbalanced within participants. Children who knew only six verbs (instead of eight) were tested on the same number of singular (three) and plural (three) videos. The last two videos of the session (one singular, one plural) were displayed on the same screen.

Coding and Reliability

A digital video camera placed between the two TV screens was used to record the child's looks during the baseline and test phases. The videos were coded offline frame by frame using SuperCoder (Hollich, 2005). For each frame, a trained coder blind to the condition (i.e., whether the audio stimuli were the singular or the plural) coded whether the child was looking at the matching video, at the nonmatching video, or away. Then, for each trial, we calculated the percentage of time that the child spent looking at the matching video compared to the nonmatching video. Trials in which children only looked at one video during baseline were discarded from the analysis. Overall, we analyzed a mean of 5.25 trials (out of a mean of 7.30 trials actually presented) per 24-month-old and a mean of 5.15 trials (out of a mean of 7.45 trials actually presented) per 30-month-old. Ten percent of the data were recoded by a different trained coder, with an average agreement of 98% for the 24-month-olds, and 94% for the 30-month-olds.

Results and Discussion

For both the baseline and test phases, the percentages of looking toward the matching videos were averaged over trials separately for trials in which the singular video versus the plural video was the target, leading to four measures per child.

Mean percentages to the matching videos were entered into a three-way analysis of variance (ANOVA) with the main between-subject factors of age (24 vs. 30 months) and the main within-subject factors of phase (baseline vs. test) and number (singular vs. plural target). The effect of phase approached significance, $F(1, 38) = 3.73$, $p = .061$,

$\eta_p^2 = .09$, indicating that the children tended to look longer at the matching videos during test ($M = 52.27\%$, $SD = 17.36$) than during baseline ($M = 47.01\%$, $SD = 8.80$). However, the triple interaction also approached significance, $F(1, 38) = 3.33$, $p = .076$, $\eta_p^2 = .08$, suggesting that the effect of phase was modulated by age and/or number. No other effect or interaction reached significance.

In order to specify the triple interaction, separate ANOVAs with the two factors of phase and number were conducted for each age group. At 24 months of age (Figure 2), both main effects were nonsignificant (both *F*s < 1). For the phase factor, only ten of the twenty 24-month-olds increased their looking times toward the target video between familiarization and test. However, the Phase × Number interaction was significant, $F(1, 19) = 7.06$, $p = .016$, $\eta_p^2 = .27$. Planned comparisons, though not reaching significance, showed that the 24-month-olds tended to decrease their looking towards the singular video in the singular target trials, $F(1, 19) = 1.29$, $p = .27$, $\eta_p^2 = .06$, while they tended to increase their looking toward the plural video in the plural target trials, $F(1, 19) = 2.96$, $p = .10$, $\eta_p^2 = .14$. Overall, this pattern shows that the 24-month-olds tended to increase their looking toward the plural video between baseline ($M = 46.37\%$, $SD = 11.68$) and test ($M = 54.02\%$, $SD = 25.16$). This was confirmed by an ANOVA with the factors of phase and number on the variable percent looking at plural video showing a main effect of phase, $F(1, 19) = 7.06$, $p = .016$, $\eta_p^2 = .27$, and no other effect or interaction, $F(1, 19) < 1$. Although it is unclear why they did so, this pattern of results brings no evidence in support of an understanding of number agreement at 24 months of age.

At 30 months of age (Figure 3), there was a main effect of phase, $F(1, 19) = 9.09$, $p = .007$, $\eta_p^2 = .32$, establishing that the older children looked longer at the matching video after having heard the speech stimulus ($M = 53.1\%$, $SD = 8.5$) than during baseline ($M = 46.3\%$, $SD = 12.9$). This conclusion is also supported by the fact that fourteen of the twenty 30-month-olds increased their looking times toward the target video between familiarization and test ($p = .058$, binomial test). There was also a main effect of number, $F(1, 19) = 5.77$, $p = .027$, $\eta_p^2 = .23$, indicating that the older children tended to look longer toward the targets in the trials of the singular condition. Importantly though, there was no interaction between the two factors, $F(1, 19) < 1$, indicating that the older children increased their looking times toward the matching video between baseline and test in both the singular and plural conditions. This pattern of results is compatible with an understanding of number agreement at 30 months of age.

The present results reveal that the linguistic stimulus did alter the older children's initial preference. While the 24-month-olds looked longer at the video depicting two boys performing an action in the testing phase, regardless of singular versus plural verbal stimulus, the 30-month-olds systematically looked longer at the matching video in the testing phase both in the singular and plural conditions, compared with the silent phase.

Because the agreement form tested in the present study relies on the contrast between third-person singular (*il*) and plural (*ils*) subject pronouns, we

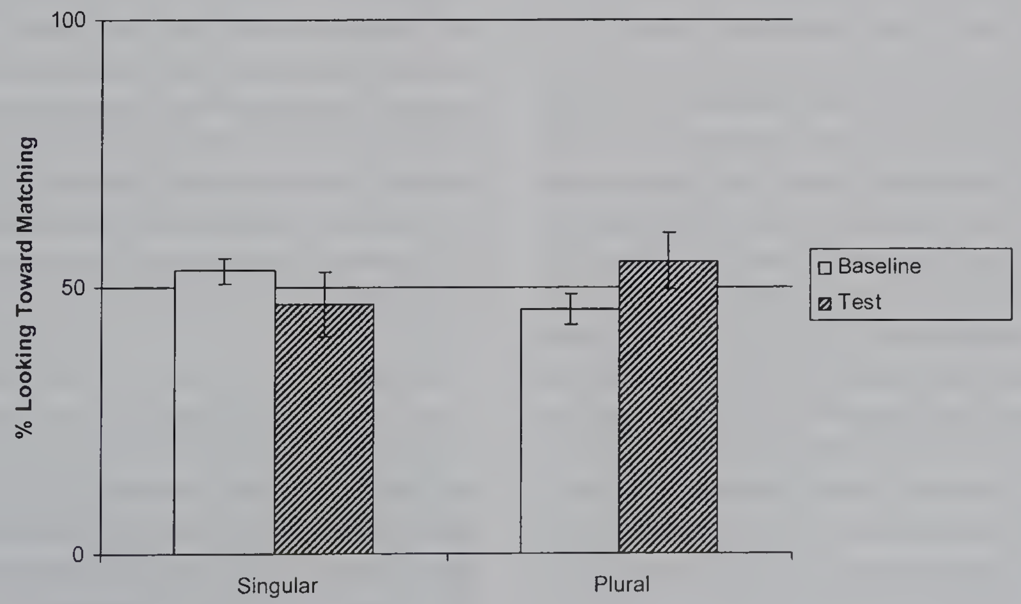


Figure 2. Mean percent looking times (and *SE*s) toward the matching video in silent baseline and at test, broken down by the number of the target, at 24 months of age.

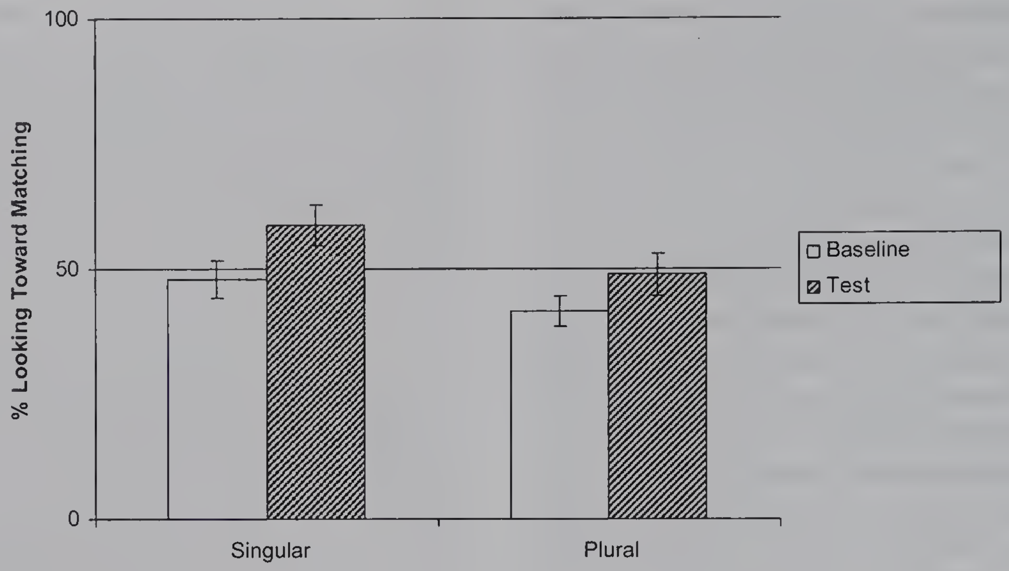


Figure 3. Mean percent looking times (and *SE*s) toward the matching video in silent baseline and at test, broken down by the number of the target, at 30 months of age.

computed children's knowledge of these elements according to parents' responses on the MCDI questionnaires filled out at the time of testing. At 24 months, 50% of the children were reported to understand singular *il* versus 29% for plural *ils*. At 30 months, 96% of the children were reported to understand singular *il* versus 88% for plural *ils*. These numbers show, first, an increase in comprehension of these pronouns between the two ages, for both *il*, $\chi^2(1, N = 35) = 6.32$, $p = .012$, $\phi = .49$, and *ils*, $\chi^2(1, N = 35) = 10.70$, $p = .001$, $\phi = .61$, suggesting that 30-month-olds may have representations of the singular and plural forms they use while 24-month-olds may not. Second, they establish a lack of singular–plural asymmetry, for both 24-month-olds, $\chi^2(1, N = 18) = 1.05$, $p = .31$, $\phi = -.23$, and 30-month-olds, $\chi^2(1, N = 17) = .34$, $p = .55$, $\phi = -.10$, corroborating the lack of an asymmetry in the IPLP results.

Correlation analyses between children's performance in the IPLP task (three variables: overall score, score on singular trials, score on plural trials) and MCDI data (six variables: total number of words/verbs/pronouns understood or produced) were also carried out for both age groups. The only significant correlation was a positive correlation found for 24-month-olds between performance on singular trials and total number of verbs understood, $r(17) = .54$, $p = .02$. No significant correlation was found at 30 months. These results suggest a rather weak link between lexical acquisition and comprehension of agreement, although caution is required in interpreting the present pattern of correlations given the small numbers of participants at each age.

Overall, the results of the present IPLP study demonstrate that 30-month-old children acquiring French are able to distinguish the singular and the plural marking expressed as the linking consonants /l/ versus /z/ in onset position of the second syllable of otherwise phonologically identical verbs. In particular, they were able to match these forms to the target stimuli, in the absence of any additional morphosyntactic or lexical cues. No such evidence could be found for 24-month-olds. These results, therefore, provide evidence for comprehension of a single number contrast by 30 months, which is probably emerging somewhere between 24 and 30 months of age. Like in Kouider et al. (2006) and Wood et al. (2009), young children showed comprehension of *verbal* number marking but there were significant differences in the stimuli. In the English case, the verbal contrast was lexical (*is* vs. *are*) and 24-month-olds showed comprehension with familiar or nonce objects only when several number cues were presented (e.g., "Look, there **are some** blickets" vs. "Look, there **is a** blicket"). When a single nominal cue in the form of the bound morpheme -*s* was presented, children younger than 36 months did not show comprehension.

Because the results of the French IPLP experiment clearly challenge the conclusion based on English and Spanish that subject–verb agreement is mastered late, around age 5 (Johnson et al., 2005; Pérez-Leroux, 2005), we set out to replicate them, using a different experimental technique. In Experiment 2, we administered a pointing task to a separate group of 30-month-olds and asked whether the 6.8% increase in looking time observed between baseline and test in IPLP revealed an understanding of number agreement stable enough to be able to guide a pointing response, a task requiring both decisional and motor planning.

Experiment 2

IPLP has been widely and successfully used to investigate other aspects of language acquisition, in particular, argument structure in younger children (Hirsh-Pasek, Golinkoff, & Naigles, 1996; Naigles, 1990, 1996; Naigles & Bavin, 2002) because it places very few demands on them. For the present agreement study, it is worth asking whether adding some burden on children would result or not in comprehension failure. This, in turn, would reveal how robust the underlying representation of subject–verb agreement is, say, at 30 months of age. Therefore, Experiment 1 was replicated for 30-month-olds using a pointing task (more similar than IPLP to the methods used in previous studies on subject–verb agreement). If their representation of subject–verb agreement is stable, pointing is likely not to hinder comprehension, especially since children in that age range are known to succeed in this task (Brandone et al., 2007; Dittmar et al., 2008; Maguire et al., 2008).

Method

Participants

Participants included 16 monolingual French-speaking 30-month-olds (8 males, 8 females; *M* = 30 months, 12 days; range = 28 months, 18 days to 32 months, 11 days). Six additional children were tested but excluded from the experiment due to their refusing to participate (1) or pointing in fewer than four trials (5). Children were all tested in Paris. They were recruited from a database of parents who volunteered to participate in child development studies and came from diverse socioeconomic backgrounds.

Materials

We used the same visual and verbal stimuli as in Experiment 1 (of the two pronunciations of the plural pronoun, /iz/ and /ilz/, the experimenter consistently used the latter). Children were only presented with verbs that they comprehended according to parental report.

Procedure

Prior to coming to the lab, parents were asked to name which of the 10 verbs used in the experiment their child knew. Right before their visit, parents also completed a French version of the MCDI modified by the authors for present purposes. MCDI questionnaires were returned for 12 (of 16) children.

Each child was tested individually in a soundproof booth for about 10 min using a pointing task. Children sat on a parent's lap during the whole session. The parent, wearing dark glasses rendered opaque by covering the inside of the lenses, was blind to the video stimuli. The (female) experimenter sat to the immediate right of, and slightly behind, the child. Having the experimenter sit fully behind the child was tested but proved problematic: The child would frequently turn back because of interest or to seek approval. When pronouncing the target utterance, the experimenter was careful not to look at the child or the screens; rather she stared at the central light between the two screens so as not to influence the child's pointing.

The session started with up to four training trials intended to set up the pointing game. The structure of the trials was similar to that in Experiment 1, and it was the same for the training and test trials. The child's attention was first centered by flashing a central light located between the two presentation screens (the child facing the whole display). Once the child looked at the central light, a pair of pictures of bright, easily recognizable objects was shown in silence for 6 s. Then the screens turned black and the experimenter named one of the two objects and asked the child to point to it (e.g., *Tu as vu la pomme? Montre-moi avec ton doigt où elle se trouve la pomme, montre-moi la pomme* "Did you see the apple? Show me with your finger where the apple is, show me the apple"). The two images immediately reappeared on their respective screens for 6 s. The child's task was to point to the matching screen. If the child didn't point during the first seconds, the experimenter repeated the invitation once to give the child another chance at responding and then, regardless of response, moved onto the next trial. Each trial ended with the 3-s eye-catching video displayed on the matching screen, as done in Experiment 1, following Kouider et al. (2006).

Each child was then presented with eight trials of familiar verbs (all knew at least eight of the verbs used in the present experiment). They were simultaneously presented the 6-s singular/plural versions of the actions used in Experiment 2 twice, first in silence then following the experimenter's request (e.g., *Tu as vu? Ils attrapent le zappe, montre-moi avec ton doigt où ils attrapent le zappe, montre moi où ils attrapent le zappe* "Did you see? They are catching the zap, show me with your finger where they are catching the zap, show me where they are catching the zap").

For each child, the target was the singular action four times and the plural action the other four times (in random order). The order of presentation of the different actions/verbs was randomized by the experimenter, and the side of the matching video was counterbalanced within participants for each condition (singular/plural).

Coding and Reliability

Coding of children' pointing was done online by the experimenter, but the sessions were videotaped using a digital video camera placed between the two screens. Each trial was coded as correct (score = 1) or incorrect (score = 0). A response was coded as correct only if the child clearly and decisively pointed to one screen. Any hesitant, double pointing or refusal to point resulted in the elimination of the corresponding trials. Overall, children clearly pointed on 7.25 of 8 trials. The data were first coded online, then entirely recoded offline by an assistant not familiar with the study and blind to the conditions of each trial. Interrater reliability was of 96%. Only five trials led to different answers by the two coders; these trials were viewed again by the two original coders and a third coder, and coding agreement was easily reached.

Results and Discussion

There was no significant difference between the singular and plural conditions, $t(15) = .48$, $p = .64$, two-tailed, $d = .17$ (Figure 4). Overall, children pointed at the matching video 61.93% ($SD = 14.16$) of the time ($M = 64.06\%$, $SD = 26.48$ for singular; $M = 59.90\%$, $SD = 23.02$ for plural, the slightly better performance for singular being entirely due to 1 child who only pointed to singular actions), which is significantly above the 50% chance level, $t(15) = 3.37$; $p = .004$, two-tailed, $d = 1.74$. There was a majority of 12 children who gave more correct points versus 2 children who gave fewer correct points and 2 children who gave as many correct and incorrect points, $\chi^2(2, N = 16) = 10.38$, $p = .006$, $\phi = .38$.

Just as in Experiment 1, the lack of singular versus plural asymmetry revealed in the pointing task results is corroborated by MCDI evidence that 30-month-olds have the same level of comprehension of singular *il* and plural *ils*: Ninety percent of the 30-month-olds in Experiment 2 (compared to 96% in Experiment 1) were reported to understand *il* versus 80% for *ils* (compared with 88% in Experiment 1). Note that both groups of 30-month-olds are clearly at the same level of acquisition of these third-person subject pronouns, both for *il*, $\chi^2(1, N = 27) = .16$, $p = .69$, $\phi = -.08$, and *ils*, $\chi^2(1, N = 27) = .34$, $p = .56$, $\phi = -.11$.

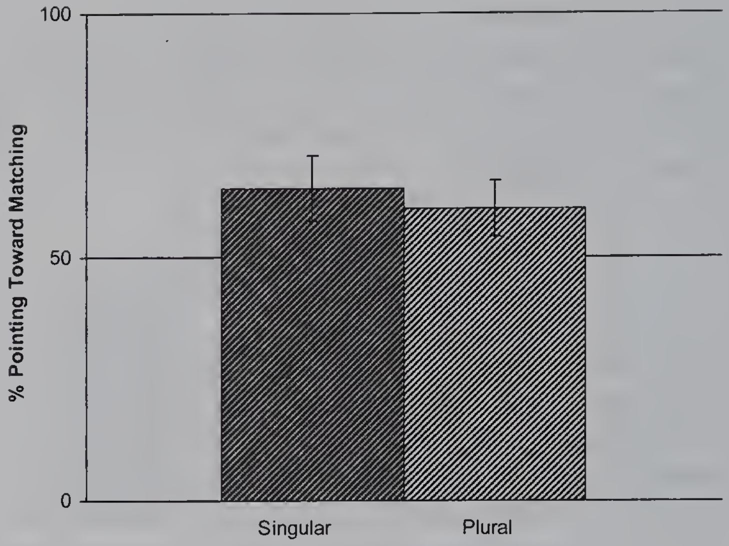


Figure 4. Mean percent pointing (and *SE*s) toward the matching video at test, broken down by the number of the target, at 30 months of age.

As in Experiment 1, we also performed correlation analyses between children's performance (three variables: overall score, score on singular trials, score on plural trials) and MCDI data (six variables: total number of words/verbs/pronouns understood or produced). There was a significant positive correlation between overall performance and total number of verbs understood, $r(11) = .59$, $p = .04$, and a marginal positive correlation between performance on singular trials and total number of verbs understood, $r(11) = .55$, $p = .06$. Again, these results suggest a rather weak link between lexical acquisition and comprehension of agreement, although caution is required in interpreting the present data given the small numbers of participants.

Overall, Experiment 2 shows that increasing the decisional and motor burden on 30-month-olds does not hinder children's performance. Thirty-month-old children were able to distinguish between third-person singular and plural and match the plural and singular expressions of liaison to the target visual stimuli. Moreover, no asymmetry between children's performance in the singular versus plural conditions was revealed by the analyses. Representations of subject–verb agreement for the forms tested, thus, appear to be rather stable at 30 months of age.

The results of Experiments 1 and 2 raise the issue of whether children's performance is closely or loosely tied to properties of the adult input. Recall that Childers et al. (2001) suggested that Spanish-learning children's preference for singular agreement forms in a pointing task was paralleled by a singular bias in the adult input, albeit in different varieties of Spanish. Does a similar pattern extend to French or not? The experimental results in Experiment 1 and 2 show no singular–plural asymmetry. Is such an asymmetry also absent in French adult input? We analyzed large samples in Experiment 3 to determine the frequencies of singular and plural bare pronoun + verb combinations similar to those used in our stimuli and tested the hypothesis put forward by Childers et al. (2001).

Experiment 3

We analyzed the occurrence in child-directed speech and in spontaneous child production of the type of third-person subject pronoun + verb combinations we used in our experimental stimuli.

Method

For our analysis of the frequency of use of the third-person singular and plural subject pronouns in adult input, we used the Child Language Data Exchange System (MacWhinney, 2000) data of spontaneous speech by native French mothers of 5 monolingual, normally developing children acquiring (Parisian) French for a total of more than 54,000 utterances: dizygotic twins Camille and Pierre, ages 15–25 months and 16–27 months, respectively (Hunkeler, 2005); Pauline, age 14–30 months (Bassano & Mendes-Maillochon, 1994); Anne, age 22–30 months (Plunkett, 2002); and Grégoire, age 21–29 months (Champaud corpus).

We coded all occurrences (full and reduced) of the third-person singular pronoun *il* (/il/ or /i/) and third-person plural pronoun *ils* (/il/, /ilz/, /iz/) in the absence of a full NP subject, distinguishing eight types of verb forms, depending on: (a) number (singular or plural), (b) whether the verb starts with a vowel or a consonant (which determines the actual pronunciation of the pronoun), and (c) whether the verb has a regular ending or form (Class I) or irregular (other classes).

If the verb is regular and starts with a vowel (Type A), marking of number is carried by the preverbal pronoun only (*Il embrasse* /ilãbras/ "He kisses" vs. *Ils embrassent* /i(l)zãbras/ "They kiss"). This is the combination type used in Experiments 1 and 2. Three other types are found in French. If the verb is regular and starts with a consonant (Type B) number distinctions are all silent; the default form /i(l)/ is used (e.g., *Il(s) danse(nt)* /i(l)dãs/ "He or they dance"). If the verb is irregular and starts with a vowel (Type C), number is redundantly marked by the pronoun and the verb root or ending (e.g., *Il a* /ila/ "He has" vs. *Ils ont* /i(l)zõ/ "They have"). If the verb is irregular and starts with a consonant (Type D), marking of number is carried by the ending or root only, the distinction between *il* and *ils* being silent again (e.g., *Il prend* /i(l)prã/ "He takes" vs. *Ils prennent* /i(l)prɛn/ "They take").

Results and Discussion

The results are displayed in Table 1 with some specific points of comparison of the adult input with the stimuli used in Experiments 1 and 2.

The percentage of child-directed utterances containing any bare third-person pronoun + verb combination out of the total number of child-directed utterances varies across children (from 4% to 18%) but it remains overall quite low. The percentage of singular combinations out of all third-person pronoun + verb combinations is always higher than plural combinations, ranging from 62% for Pierre to 92% for Grégoire and disproportionally so for four of five children. Overall, this means that French-learning children are not systematically exposed to a large proportion of third-person pronoun + verb combinations compared to other types of utterances (e.g., noun + verb, first- or second-person subject pronoun + verb). Nor are they systematically exposed to any morphophonological contrast between third-person singular and plural pronoun + verb combinations, regardless of verbal type given the marked asymmetry in favor of singular forms.

Type A forms (in bold) are similar to the forms used in our comprehension stimuli. As shown in Table 1, Type A forms are relatively rare in the adult input (except for Camille, 39%) compared to other types of verb forms, both in the singular (2%–8%, out of all singular combinations) and in the plural (0%–12%, out of all plural combinations).

A parallel analysis of spontaneous child production in the same corpora is displayed in Table 2. Up to 30 months of age, children vary in their overall level of spontaneous production of bare third-person pronoun + verb combinations. Yet

Table 1
Frequency of Adult Third-Person Pronoun + Verb Combinations in Adult Input (Type A = Combinations Used in Experimental Stimuli)

Corpus and age in months	Camille (15–25)	Pierre (16–27)	Pauline (14–30)	Anne (22–30)	Grégoire (21–29)
Total adult utterances	1,875	1,558	8,873	35,480	6,430
Total singular and plural combinations, Types A–D (out of all adult utterances)	8% (151/1,875)	18% (279/1,558)	5% (421/8,873)	4% (1,374/35,480)	10% (639/6,430)
Singular combinations, Types A–D (out of all combinations)	91% (137/151)	62% (174/279)	90% (381/421)	82% (1,130/1,374)	92% (589/639)
Singular combinations, Type A (out of all singular combinations)	39% (53/137)	2% (3/174)	8% (30/381)	7% (79/1,130)	8% (45/589)
Plural combinations, Type A (out of all plural combinations)	0% (0/14)	1% (1/105)	5% (2/40)	6% (14/244)	12% (6/50)

Table 2
Frequency of Spontaneous Child Third-Person Pronoun + Verb Combinations (Type A = Combinations Used in Experimental Stimuli)

Corpus and age in months	Camille (15–25)	Pierre (16–27)	Pauline (14–30)	Anne (22–30)	Grégoire (21–29)
Total child utterances	1,334	1,057	3,943	5,468	4,037
Total singular and plural combinations, Types A–D (out of all adult utterances)	0% (0/1,334)	0% (0/1,057)	1% (46/3,943)	3% (150/5,468)	3% (133/4,037)
Singular combinations, Types A–D (out of all combinations)	0% (0/0)	0% (0/0)	89% (41/46)	83% (125/150)	94% (125/133)
Singular combinations, Type A (out of all singular combinations)	0% (0/0)	0% (0/0)	0% (0/41)	2% (3/125)	3% (4/125)
Plural combinations, Type A (out of all plural combinations)	0% (0/0)	0% (0/0)	0% (0/5)	0% (0/25)	0% (0/8)

they are systematically found to produce more singular than plural forms if they produce any (ranging from 83% to 94%), and very few vowel-initial verbs.

In sum, the analysis of five corpora provides a clear picture of the particular verbal forms of the language French-learning children are generally exposed to. First, they are exposed to a much higher proportion of singular than plural third-person pronoun + verb combinations across all types of morphophonological instantiations. Second, vowel-initial verbs varying in the singular versus plural because of liaison constitute an infrequent morphophonological subclass in the adult input. Therefore, French-learning children are rarely exposed to the type of stimuli we used in comprehension Experiments 1 and 2, a situation different from the relatively frequent subject–verb agreement forms studied in Spanish, and to a lesser extent, English. Moreover, like its Spanish counterpart, the French adult input displays a singular bias; however, such a bias was not reflected in our comprehension results, contrary to the singular bias found in Spanish by Childers et al. (2001).

General Discussion

The main finding of the present study is that French-learning 30-month-olds distinguish singular versus plural subject–verb agreement forms signaled in the plural by a /z/ sound at the onset of the second syllable of a simple verbal utterance and use them to guide looking or pointing to events with singular or plural agents. While 24-month-olds did not succeed at the preferential looking task, 30-month-olds were able to match the same visual scene with a corresponding verbal stimulus in two different experimental tasks, with and without added decisional and motor planning with no incidence on the results (the often reported increased sensitivity of looking over pointing tasks is thus not found in our study). This suggests that 30-month-olds have knowledge of subject–verb agreement, and that this knowledge is independent of the demands of the task and therefore robust.

While early sensitivity to number agreement cues may be present in English-learning children as young as 16 months (Soderstrom et al., 2007) and comprehension of singular–plural morphology has been found in 24-month-olds provided it is lexicalized and occurs in combination with nominal cues (Kouider et al., 2006; Wood et al., 2009), comprehension of bound singular–plural morphology had not been found before the age of 3 or 5, depending on the study (Childers et al., 2001; cf. Johnson et al., 2005; Pérez-Leroux, 2005). To the best of our knowledge the results of Experiments 1 and 2 are the first to date that point to an earlier age—30 months—for comprehension of bound number morphology in the verbal domain in the absence of any additional cue.

Why do we find such a difference in age of comprehension of subject–verb agreement in French versus English and Spanish? Our main hypothesis relates it to the different encoding of number in the languages investigated: suffixal in English and Spanish versus "prefixal" in French. In some speech perception and word recognition/learning tasks, children have been shown to process consonants better in onset position than in coda position (Swingley, 2003; Zamuner, 2006; though see Nazzi & Bertoncini, 2009; Swingley, 2009, for findings failing to find an onset/coda asymmetry). Accordingly, if the /z/ consonant that marks plural agreement in French through the phenomenon of liaison was treated as an onset consonant (a likely possibility that will have to be directly investigated in the future), it might explain why we found better/earlier processing than for the –/s/ and –/n/ coda consonants marking agreement in English and Spanish, respectively. It is as if agreement marking earlier in the word entails that it is heard precisely at the time children are considering an interpretation of the verb (as onset consonants seem to have more influence on processing in early word recognition tasks; Swingley, 2003).

This difference in encoding may not, however, explain the size of the age gap (30 months/2 years and 6 months vs. 5 years of age). Potentially crucial procedural differences, in particular in the nature and quality of the stimuli used, may also have contributed to the results. In both the English and Spanish studies conducted by Johnson et al. (2005) and Pérez-Leroux (2005), children looked at static pictures of, say, one duck versus two ducks swimming. The pictorial representation of the action was necessarily abstract, with only a wavy line depicting water, and no representation of swimming per se. This may have resulted in insufficient salience of the verb causing subjects to focus on the subject for number information that was by design masked by using *s*-initial verbs (the stimuli used in Childers et al. (2001) appear to have been similar but full details of illustration are lacking). Using dynamic videos in the French study with actual children performing various actions may have had a facilitating effect.

However, it is also possible that the use of high-quality images rather than simplified, abstract pictures, and not the dynamic dimension of our stimuli per se, helped the 30-month-olds in our study. It could even be that the dynamic actions added some processing weight, as the discriminating elements between the singular and plural versions of the actions were only observable during a limited portion of the video. Then maybe the failure of the 24-month-olds is not due to the lack of abstract representations of subject–verb agreement but rather to difficulties at coprocessing the complex visually dynamic stimuli and the verbal stimulus within 6 s. If so, using high quality images extracted from actual videos, such as a representative frame of the singular and plural conditions, may achieve the goal of providing simpler stimuli while retaining the saliency of the action denoted by the verb and help 24-month-olds show comprehension of subject–verb agreement. On the contrary, should 24-month-olds fail to show comprehension under such conditions, then we would be closer to concluding that abstract representations of subject–verb agreement emerge somewhere between 24 and 30 months of age. In light of the results obtained by Kouider et al. (2006) and Wood et al. (2009), this would suggest that comprehension failure in 24-month-olds is tied instead to the nonredundant and nonlexical marking of number agreement contrasts in French.

The present findings have two implications for the acquisition of subject–verb agreement. First, comparing her Spanish results with English results by Johnson et al. (2005), Pérez-Leroux (2005) concluded that children show a preference for the number category that is instantiated by overt morphology, namely, singular *–s* in English and plural *–n* in Spanish. This conclusion is challenged by our results. Children failed to show better performance on plural verbal forms than singular ones despite the fact that plural agreement is always signaled by /z/ with vowel-initial verbs in French. Second, Childers et al. (2001) attributed their finding that young Spanish-learning children show a significant preference for third-person singular verbal forms over plural forms to a disproportionate number of singular forms in the adult Spanish input. This interpretation is not supported by our results, as the French-learning children we tested did not display an overall preference for singular verbal forms, despite a singular bias in the adult input.

Our findings also bear on previous claims concerning the acquisition of liaison in French because the plural number agreement tested is phonologically realized through liaison. In a picture description task requiring children to produce pronoun + vowel-initial verb combinations like *Ils écrivent* "They write" after the prompt *Ici les garçons lisent et là?* "Here the boys read and there?" Gallot, Spinelli, and Chevrot (2008) showed that 31- to 46-month-olds produce correct, obligatory liaison only 68% of the time. Experiments 1 and 2 provide evidence for awareness of liaison in verbal contexts by 30 months, an age when children's production of liaison is not yet systematic. A possible explanation for this discrepancy relates to Gallot et al. (2008)'s use of a picture description task requiring children to cooperate and remember a prompt. This explanation is indirectly supported by independent studies having revealed a gap between production of *–s* marking in English in the context of a sentence completion with nonce verbs task (Berko, 1958) versus spontaneous production (Brown, 1973).

Our study also contributes to the growing literature on the role played by function words in acquiring language. Previous studies established that subject pronouns like *il* facilitate early segmentation of verbs and comprehension of verbs as denoting actions. Regarding segmentation, Marquis and Shi (2008) tested French-learning 11-month-olds on novel verb segmentation on passages containing two combinations of subject noun + verb and five combinations of a subject pronoun + verb, and found that these children at the onset of vocabulary learning were able to segment monosyllabic verbs out of two-word sequences. Regarding comprehension, Bernal et al. (2007), using a pointing task, taught novel verbs to 23-month-old French children watching a moving object within sentences consisting of the subject pronoun *il* + verb (e.g., *Il poune par là* "It's pooning there") and novel nouns to a control group (e.g., *Un poune est là* "A poone is here") with the same visual stimuli. These children were able to distinguish verb/actions from noun/objects on the basis of their syntactic environment, the pronoun *il* versus the determiner *un*. Our study adds to this body of research on function words by establishing that by 30 months of age, children comprehend number subject–verb agreement based on pronoun-related phonological form variations.

Finally, our study has implications for general theories of syntactic acquisition. One of the main current debates in first language acquisition focuses on whether early acquisition proceeds via item-based schemas closely tied to input frequency (e.g., Lieven, Pine, & Baldwin, 1997; Theakston & Lieven, 2005; Tomasello, 2000) or abstract representations largely independent of input frequency (e.g., Chomsky, 1981; Fisher, 2002; Lidz & Gleitman, 2004). Overall, the results of the present experimental studies combined with the corpus study challenge attempts at establishing a close connection between child performance and adult input, as children were clearly sensitive to both singular and plural agreement relations at an early age despite a pattern of infrequent and singular-dominant use of these forms by adults.

Although constructivist accounts do not make precise claims about the amount of input required for item-based learning of singular–plural marking, our results nevertheless indicate that very little input is needed to master the liaison-based subject–verb agreement system of French. Accordingly, we interpret our data as evidence that 30-month-old French-learning children have reached a milestone in the acquisition of subject–verb agreement in that they have acquired the abstract categories of singular/plural instantiated by a single cue, a third-person singular/plural pronoun triggering liaison, as well as the contrast between singular and plural. Their linguistic representations of number are abstract in kind rather than item-based, as well as robust. In a follow-up study, we plan to test whether these representations are robust enough to support comprehension in noun/verb variants of Experiments 1 and 2 in 30-month-olds.

To end on this issue, because language acquisition is not possible without linguistic input, the present results raise the question of how 30-month-olds have learned subject–verb agreement given the low frequency of the relevant forms in the adult input. We suspect that it is the cumulative effect of acquiring the entire system of subject–verb agreement in French that makes it possible to learn all its expressions, ranging from the morphological marking of number on the subject NP, doubling by a pronoun, the verb itself if irregular, as well as liaison where appropriate. A particularity of spoken French including child-directed French (Culbertson & Legendre, 2008) may be relevant. Spoken French favors pronoun doubling or the co-occurrence of a subject NP and a pronoun agreeing in person and number (e.g., *Jean il est parti* "John he left"; *Marie elle est partie* "Mary she left"). Depending on whether the verb itself displays regular or irregular person/number marking, whether the verb is vowel or consonant initial, and whether the subject NP includes a determiner or not, a nonadjacent dependency results that may include up to four cues: ***L'animal il a*** *une trompe* "The animal it has a trunk" versus ***Les animaux ils ont*** *une trompe* "The animals they have a trunk." The redundancy in morphophonological cues of number including two identically sounding liaisons marking plural (*Les animaux* /le**z**animo/; *ils ont* /i(l)**z**õ/) may well be salient enough for abstract representations to follow and for the use of a single cue with low saliency in the interpretation of third-person singular versus plural subjects by 30-month-old children.

References

Anisfeld, M., & Tucker, R. (1967). English pluralization rules of six-year-old children. *Child Development*, *38*, 1201–1217.

Auger, J. (1994). *Pronominal pronouns in Québec Colloquial French: A morphological analysis* (Doctoral dissertation). University of Pennsylvania, Philadelphia.

Bassano, D., & Mendes-Maillochon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: A longitudinal study. *Journal of Child Language*, *21*, 649–675.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb learning. *Language Learning and Development*, *3*, 325–341.

Brandone, A. C., Pence, K. L., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Action speaks louder than words: Children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development*, *78*, 1322–1342.

Brown, R. W. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Childers, J., Fernandez, A. M., Echols, C., & Tomasello, M. (2001). Experimental investigations of children's understanding and use of verb morphology: Spanish- and English-speaking 2 ½ and 3-year-old children. In M. Almgen, A. Bareña, M. J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on language acquisition. Proceedings of the International Association for the Study of Child Language* (pp. 104–127). Somerville, MA: Cascadilla Press.

Chomsky, N. (1981). *Lectures on Government and Binding*, Dordrecht, Netherlands: Foris.

Culbertson, J., & Legendre, G. (2008). Qu'en est-il des clitiques sujet en français oral contemporain? In J. Durand, B. Habert & B. Laks (Eds.), *Actes du 1er Congrès Mondial de Linguistique Française* (pp. 2651–2662). Paris: EDP Sciences.

De Villiers, J., & de Villiers, P. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*, 267–278.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, *79*, 1152–1167.

Encrevé, P. (1988). *La liaison avec et sans enchaînement, phonologie tridimensionnelle et usage du français.* Paris: Editions du Seuil.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* San Diego, CA: Singular.

Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, *5*, 55–64.

Fonseca-Greber, B., & Waugh, L. (2003). The subject pronouns of European conversational French: Morphologization, grammatical change, semantic change and change in progress. In R. Nùñez-Cedeño & L. López (Eds.), *Romance perspective on language knowledge and use* (pp. 99–118). Amsterdam: John Benjamins.

Gallot, S., Spinelli, E., & Chevrot, J.-P. (2008). Acquisition of liaison: The test of the pronoun-verb sequences. Paper presented at the Congress of the International Association for the Study of Child Language (IASCL08), Edinburgh, UK.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23–45.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar.* Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Golinkoff, R. M., & Naigles, L. R. (1996). Young children's use of syntactic frames to derive meaning. In K. Hirsch-Pasek & R. Golinkoff (Eds.), *The origins of grammar* (pp. 123–158). Cambridge, MA: MIT Press.

Höhle, B., & Weissenborn, J. (2003). German-learning children's ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*, 122–127.

Hollich, G. (2005). Supercoder: A program for coding preferential looking (Version 1.5) [computer software]. West Lafayette, IN: Purdue University.

Hunkeler, H. (2005). *Aspects of the evolution of the interactions mother-child: Case study of two dizygotic twin children between 15 and 26 months* (Master's thesis), University of Rouen, France.

Jakubowicz, C., & Rigaut, C. (1997). L'acquisition des clitiques nominatifs en français. In A. Zribi-Hertz (Ed.), *Les pronoms: Morphologie, syntaxe et typologie* (pp. 57–99). Saint-Denis, France: Presses Universitaires de Vincennes.

Johnson, V. A., de Villiers, J. G., & Seymour, H. N. (2005). Agreement without understanding? The case of third person singular /s/. *First Language*, *25*, 317–330.

Kern, S. (2003). Le compte-rendu parental au service de l'évaluation de la production lexicale des enfants français entre 16 et 30 mois. *Glossa*, *85*, 48–62.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*, 1–25.

Lambrecht, K. (1981). *Topic, antitopic and verb agreement in non-standard French. Vol. II: 6 of Pragmatics and beyond.* Amsterdam: John Benjamins.

Le Normand, M. T. (1986). A developmental exploration exploration of langauge used to accompany symbolic play in young normal children (2–4 years old). *Child Care, Health and Development*, *12*, 121–134.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavioral Research Methods, Instruments & Computers*, *36*, 156–166.

Lidz, J., & Gleitman, L. (2004). Yes, we need Universal Grammar. *Cognition*, *94*, 85–93.

Lidz, J., Waxman, S., & Freedman, J. (2003). What children know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18-months. *Cognition*, *89*, 65–73.

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically based learning and early grammatical development. *Journal of Child Language*, *24*, 187–217.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* (3rd ed.). Mahwah, NJ: Erlbaum.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*, 628–634.

Marquis, A., & Shi, R. (2008). Segmentation of verb forms in preverbal infants. *Journal of the Acoustical Society of America*, *123*(4), EL105–EL110.

Miller, P. (1992). *Pronouns and constituents in phrase structure grammar*. New York: Garland.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*, 357–374.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*, 221–251.

Naigles, L., & Bavin, E. (2002). Generalizing novel verbs to different structures: Evidence for the early distinction of verbs and frames. In B. Skarabela, S. Fish, & A. H. J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 417–428). Somerville, MA: Cascadilla Press.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*, 13–30.

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech*, *52*, 463–480.

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*, 522–537.

Nazzi, T., & Pilardeau, M. (2007). Learning the names of objects is not enough to support name-based categorization at 16 months. *European Journal of Developmental Psychology*, *4*, 435–450.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *Année Psychologique*, *101*, 447–462.

Pérez-Leroux, A.-T. (2005). Number problems in children. In C. Gurski (Ed.), *Proceedings of the 2005 Canadian Linguistic Association Annual Conference* (p. 12). Retrieved from http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm

Pierce, A. (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht, Netherlands: Kluwer.

Plunkett, B. (2002). Null Subjects in child French interrogatives: A view from the York Corpus. In C. Pusch & W. Raible (Eds.), *RomanistischeKorpuslinguistik: Korpora und gesprochene Sprache / Romance Corpus Linguistics: Corpora and Spoken Language* (pp. 441–452). Tübingen, Germany: Gunter Narr Verlag.

Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America*, *106*, 1555–1565.

Sachs, J. (1983). There and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's language* (Vol. 4, pp. 1–28). Hillsdale, NJ: Erlbaum.

Shady, M. (1996). *Children' sensitivity to function morphemes* (Doctoral dissertation). State University of New York, Buffalo.

Shi, R., Marquis, A., & Gauthier, B. (2006a). Segmentation and representation of function words in preverbal French-learning infants. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 549–560). Somerville, MA: Cascadilla Press.

Shi, R., Werker, J. F., & Cutler, A. (2006b). Recognition and representation of function words in English-learning infants. *Infancy*, *10*, 187–198.

Soderstrom, M. (2002). *The acquisition of inflection morphology in early perceptual knowledge of syntax*. (Doctoral dissertation). Johns Hopkins University, Baltimore, MD.

Soderstrom, M. (2008). Early perception – late comprehension of grammar? The case of verbal -s: A response to de Villiers & Johnson (2007). *Journal of Child Language*, *35*, 671–676.

Soderstrom, M., Wexler, K., & Jusczyk, P. (2002). English-learning children's sensitivity to agreement morphology in receptive grammar. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (Vol. 2, pp. 643–652). Somerville, MA: Cascadilla Press.

Soderstrom, M., While, K. S., Conwell, E., & Morgan, J. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy*, *12*, 1–29.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103–114.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, *46*, 265–294.

Swingley, D. (2009). Onsets and codas in one-year-olds' word recognition. *Journal of Memory and Language*, *60*, 252–269.

Theakston, A., & Lieven, E. (2005). The acquisition of auxiliaries BE and HAVE: An elicitation study. *Journal of Child Language*, *32*, 587–616.

Tomasello, M. (2000). Do children have adult syntactic competence? *Cognition*, *74*, 209–253.

Tomasello, M., & Olguin, R. (1993). Twenty-three-month-old children have a grammatical category of noun. *Cognitive Development*, *8*, 451–464.

Tranel, B. (1996). French liaison and elision revisited: A unified account within Optimality Theory. In C. Parodi, C. Quicoli, M. Saltarelli, & M. L. Zubizarreta (Eds.), *Aspects of romance linguistics* (pp. 433–455). Washington, DC: Georgetown University Press.

Valian, V., & Aubry, S. (2005). When opportunity knocks twice: Two-year-olds' repetition of subjects. *Journal of Child Language*, *32*, 617–641.

Valian, V., Prasada, S., & Scarpa, J. (2006). Direct object predictability: Effects on children's imitation of sentences. *Journal of Child Language*, *33*, 247–269.

Wood, J. S., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology*, *45*, 202–206.

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9-to-16-month-old infants. *Infancy*, *10*, 77–95.

Zaroukian, E. (2009). *Form, meaning, and comprehension/production asymmetries in agreement* (Master's thesis). Johns Hopkins University: Baltimore, MD.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1876–1893

Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading

Serena Lecce, Silvia Zocchi,
Adriano Pagnin,
and Paola Palladino
University of Pavia, Italy

Mele Taumoepeau
University of Otago, New Zealand

The relation between children's mental state knowledge and metaknowledge about reading was examined in 2 studies. In Study 1, 196 children (mean age = 9 years) were tested for verbal ability (VA), metaknowledge about reading, and mental state words in a story task. In Study 2, the results of Study 1 were extended by using a cross-lagged design and by investigating older children (N = 71, mean ages = 10 years at Time 1 and 11 years at Time 2) for mental state knowledge, metaknowledge about reading, and VA. Results showed a significant relation between early cognitive (but not emotion) mental state knowledge and later metaknowledge about reading, controlling for VA. Results suggest close links between different aspects of children's knowledge about the mind.

Children's knowledge about the mind represents one of the most productive areas of research in child development. Extensive research has documented the developmental trajectory of preschool children's theory of mind (ToM) or the ability to attribute mental states to themselves and others. A well-established development in children's ability to explain, predict, and influence others' behaviors is the understanding that people may act on the basis of false beliefs (Premack & Woodruff, 1978; Wellman, Cross, & Watson, 2001). Despite this wealth of research, less is known about the consequences of children's earlier knowledge about mental states as it relates to aspects of older children's mind understanding such as metacognitive ability, that is, the ability to reflect on cognitive activity that takes itself as the cognitive object or that regulates any aspect of cognitive activity (Flavell, Miller, & Miller, 1993). This is partly due to the fact that research on metacognition has focused on school-aged children's knowledge of cognitive processes, in an effort to understand which variables affect children's performance on school tasks, whereas studies on mind understanding under the rubric of ToM are mainly interested in investigating processes underlying children's social behavior (Flavell, 2000).

In this article, our interest is directed toward disentangling the relation between these two components of children's knowledge about the mind by examining the relative and differential influence of children's cognitive and emotion mental states knowledge on children's metacognitive knowledge about reading comprehension (i.e., metaknowledge about reading). We were especially interested in metaknowledge about reading as it directly relates to children's ability to comprehend text (Brown, Armbruster, & Baker, 1986; Orasanu, 1986; Paris, Cross, & Lipson, 1984) and is thus crucial for children's school success (Garner, 1987; Schoonen, Hulstijn, & Bossers, 1998). Metaknowledge about reading refers to knowledge about the nature of people as readers (such as their attitudes and individual differences), the nature of reading as a cognitive task (such as its different aims depending on readers'goals), and about possible strategies that can be applied to the task, as well as knowledge about different types of texts and their specific characteristics. Metaknowledge about reading strategies could include knowledge of the beneficial aspects of strategies such as skimming, gaining an overview of a text, and rereading.

Many thanks to the children who took part in this study and their parents and school staff for their interest and support of this project. We also would like to thank Ted Ruffman, Henry Wellman, and a second anonymous reviewer for their extremely helpful comments on previous versions of this article.

Correspondence concerning this article should be addressed to Serena Lecce, Department of Psychology, University of Pavia, P.zza Botta 6, 27100 Pavia, Italy. Electronic mail may be sent to slecce@unipv.it.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0019

Furthermore, in mature readers, metaknowledge about texts is indicated in their sensitivity to text genres and text difficulties (Orasanu, 1986). Next, we discuss the three main aims of the present article.

The Relation Between Metaknowledge About Reading and Children's Mental State Knowledge

As stated above, the first aim of this study was to investigate the existence of a relation between children's mental state knowledge and metaknowledge about reading. The relation between metaknowledge about reading and children's mental state can be better understood if we think that a critical aspect of children's mental state understanding is that mental states are representations of the world and as such, can misrepresent (Perner, 1991). This ability can be extended to an understanding that text is also a representation that can be interpreted by oneself or others in various ways such that one could misrepresent the intended meaning of a text (Kamavar & Homer, 2000). Thus, the extent to which children understand text as metarepresentational likely depends to a certain extent on their understanding that a text has an intended meaning. These insights potentially act as an important basis for individual differences in children's capacity to reflect on the process of understanding text. Given these considerations, we hypothesized a relation between children's mental state knowledge and their metaknowledge about reading. This view can be also supported by two more general considerations. First, children's metaknowledge, defined as knowledge about the nature of people as cognisers and about cognitive processes, comprises and presupposes a knowledge about mental states. In this vein, Wellman argued that metacognition can be viewed "as a large, multifaceted theory of mind" (Wellman, 1985, p. 29). Second, theoretically, children's mental state knowledge and their metaknowledge can be considered as two distinct components of the same construct (Kuhn, 1999, 2000; Wellman, 1985), situated in the larger theoretical construct of metaknowing (Baker, 1994; Day, Cordon, & Kerwin, 1989; Flavell, 2000; Kuhn, 2000). Under this umbrella concept, Kuhn (2000) has further distinguished between *metacognitive knowing* or a declarative knowing, that is, "I know that *I know*" versus procedural knowledge or *metastrategic knowing*, which focuses on the impact of cognitive processes on-task performance, or "I know *how*."

The Specificity of the Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading

Our second goal was to investigate the specific type of mental state knowledge that would be related to children's metaknowledge about reading by comparing children's cognitive versus emotion mental state knowledge. To address this question, we analyzed four distinct aspects of children's mental state knowledge: cognitive state words (CSW), cognitive inference, emotion state words, and emotion understanding. We considered the first two variables (CSW and cognitive inference) as measures of children's cognitive mental state knowledge and the last two (emotion state words and emotion understanding) as indices of children's emotion mental state knowledge.

The rationale for investigating the differential link between cognitive versus emotion mental state knowledge and metaknowledge about reading comes from a growing body of research showing that (a) when language and background variables are taken into account, the association between children's cognitive and emotion understanding (Cassidy, Werner, Rourke, Zubernis, & Balaraman, 2003; Cutting & Dunn, 1999) and between cognitive and emotion state words (ESW) (Hughes, Lecce, & Wilson, 2007) disappears, and (b) cognitive and emotion understanding show distinct antecedents, correlates and sequelae (Cutting & Dunn, 1999; Dunn, 1995; Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991; McElwain & Volling, 2002). For example, whereas early emotion understanding is related to later positive peer experience (Denham, McKinly, Couchoud, & Holt, 1990), early cognitive understanding is associated with later negative perception of school (Dunn, 1995).

Distinguishing between cognitions and emotions is particularly important with respect to metaknowledge about reading because in order to represent their minds working with written sentences or passages, reading aloud and silently, and asking or answering questions, children should take special advantage from being familiar with cognitive mental representations and with cognitive state terms. Furthermore, children's ability to reflect on how their own minds work when reading, seems to presuppose a familiarity with terms that refer to cognitive activities (which we measured through the frequency of cognitive state talk) and to require the skill to reflect on cognitions (which we measured through cognitive inferences). Therefore, we hypothesized that children's cognitive mental state knowledge would have a stronger relation with metaknowledge about reading than their emotion mental state knowledge.

The Direction of the Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading

In line with the above proposal our last goal examined the prediction that children's mental state knowledge helps them develop their metaknowledge about reading. There are several reasons for positing a predictive relation between children's knowledge about mental states and their metaknowledge about reading. First, Kuhn (2000) proposed that moving from a knowledge of mental states in others to being able to reflect on higher order cognitive processes reflects a developmental progression that continues to develop into adulthood. We propose such a progression with respect to the development of children's metaknowledge of reading. Second, children approach the experience of learning to read at 5–6 years of age when they have already acquired a good knowledge about mental states. For example, they understand false beliefs and can distinguish mental states from real-life events (Wellman et al., 2001). The understanding that mental states can be different from reality is a critical step in children's developing knowledge about the mind and is likely to affect their metaknowledge about reading. Understanding that text can be interpreted in a variety of ways might provide insight into metastrategic knowledge, that is, that there are a variety of processes involved in understanding text. Indeed, when children come to understand that mental states and mental representations are not a simple copy of reality, they start to reflect on how their mind works in other cognitive tasks (Kuhn, 2000). Consistent with this view, preschoolers have limited awareness of the occurrence and the content of thinking (Flavell, Green, & Flavell, 1998, 2000). Finally, knowing about the cognitive processes involved in text comprehension (metaknowledge about reading) presupposes knowing about the components of that process (that the writer has an intended meaning that may or may not be grasped by the reader). Thus, having a concept of mental states provides a foundation for thinking about those states and their relation to cognitive tasks (Bartsch & Estes, 1996).

In order to address these three main goals we designed two studies. Study 1 examined the concurrent relation between metaknowledge about reading and mental state knowledge. Study 2 addressed our final two goals and is outlined in more detail after the results of Study 1. Both studies focus on school years when children's mental state knowledge becomes more sophisticated (Brown et al., 1986; Carpendale & Lewis, 2006) and metacognitive knowledge starts to develop in relation to reading comprehension. Thus, although by school age most children would have already passed first order false belief tasks, there are still strong individual differences in their tendency to refer to mental states to explain others' behaviors (Brown et al., 1986; Fox, 1991; Meins, Fernyhough, Johnson, & Lidstone, 2006; O'Connor & Hirsch, 1999). Our interest was in examining how this individual variation might relate to metaknowledge of reading.

Study 1

Study 1 addressed the first goal examining the *existence* of a relation between children's mental state knowledge and their metaknowledge about reading. In addition, we considered language and reading comprehension as potential mediators of this relation. There is a general consensus that children's verbal abilities and reading comprehension significantly correlate with their mental state knowledge (e.g., Cutting & Dunn, 1999; Hughes & Dunn, 1997; Hughes, Fujisawa, Ensor, Lecce, & Marfleet, 2006; Pelletier & Astington, 2004) and metaknowledge (Garner, 1987; Lockl & Schneider, 2007). Thus, we were interested in the relation between mental state knowledge and metaknowledge about reading over and above the effects of language and comprehension abilities. However, as stated above, we hypothesized that this would not be entirely mediated by language and reading comprehension.

In this first study we examined two age groups: 7 years and 9 years. We chose these age groups because previous research showed that they differ in their mental state knowledge (Antonietti, Liverta-Sempio, Marchetti, & Astington, 2006; Carpendale & Lewis, 2006) and metaknowledge about reading (Orasanu, 1986), with 9-year-olds showing a better understanding of mental states and a more mature knowledge of reading goals and simple strategies. We were, therefore, interested in comparing a younger and older age group in order to determine if school experience was a factor in the purported relation.

In Study 1 (as well as in Study 2), we used children's mental state words (MSW) produced in a written narrative task as a measure of their mental state knowledge. Children were asked to write an ending for two mentalizing stories concerning episodes of daily life, in which they needed to input a false belief (Story 1) and different beliefs (Story 2) in order to understand the situation described in the stories. Narratives offer an important opportunity to organize social experience (Bruner, 1986) and reflect on mental states (Dyer, Shatz, & Wellman, 2000), and as such are a useful way of investigating individual differences. At least two other studies (Camaioni, Longobardi, & Bellagamba, 1998; Fox, 1991) have adopted a written procedure asking children to produce stories that were then coded for MSW. The frequency of MSW is an important marker of individual differences in children's mental state knowledge and a good index of psychological understanding (Bartsch & Wellman, 1995; Hughes & Dunn, 1998; Hughes et al., 2007). Furthermore, as Peterson and Slaughter (2006) pointed out, the analysis of MSW in narratives allows the "exploration of individual differences in the use children make of their ToM capacities in populations from whom the false belief test is no longer informative" (p. 174).

Children's metaknowledge about reading was assessed by a questionnaire largely and successfully used in the Italian context to examine individual differences in metaknowledge both in typical and atypical development (see, e.g., Cornoldi, De Beni, & Pazzaglia, 1996; Kolić-Vehovec & Bajšanski, 2007; Pazzaglia, Cornoldi, & Tressoldi, 1993). For the purposes of the present study we selected age-appropriate subtests assessing children's understanding of the purpose of reading and strategies for reading (De Beni & Pazzaglia, 1994; Orasanu, 1986; Pazzaglia, De Beni, & Cristante, 1994). Both subtests reflect what Kuhn (2000) refers to as metastrategic knowing, that is, knowledge of how the mind works when performing a cognitive task.

Method

Participants

One hundred and ninety-six Italian children (92 boys and 104 girls), predominantly White, took part in this study. Of these, 90 attended second grade and 106 fourth grade in schools in the North of Italy. The schools were located in areas with mixed socioeconomic backgrounds. Children attending second grade (47 boys and 43 girls) had a mean age of 7.9 years (*SD* = 0.49, range = 6.3–10.25 years). Children attending fourth grade (44 boys and 61 girls) had a mean age of 10.1 years (*SD* = 0.40, range = 9.5–11.58).

Materials and Procedure

Children were administered a standardized test of metaknowledge about reading (reading goals and reading strategies), a mental state knowledge task (MSW), a standardized vocabulary task, and a reading comprehension task. All tasks were administered at school during two morning sessions of 30 min each. The order of the tasks was fixed. During the first session, children were administered first the vocabulary task and then the metaknowledge about reading task. In the second session, children were administered the story task first and then the reading comprehension task. Although the majority of the children had complete data, some of them did not complete all tasks. Eleven children did not complete the story task and the verbal ability (VA) task. Thirty-one did not complete the metaknowledge about reading questionnaire and the text comprehension task. Therefore, the numbers of children examined in individual analyses varied slightly. No age, $t(177) = 1.06$, *ns*, or gender, $\chi^2(1, 194) = 0.97$, *ns*, differences were found between children who completed all the tasks and those who gave missing data.

Children's metaknowledge about reading. This measure was assessed through the subtests of metaknowledge about reading goals (10 items) and reading strategies (10 items) from the Pazzaglia, De Beni, and Cristante's (1994) Metacomprehension Questionnaire. The reading goals subtest examines children's ability to understand that comprehension (and not decoding) is the main aim of reading and that it is possible to read for different reasons (e.g., for pleasure, for studying, for seeking specific information). The reading strategies subtest assesses children's knowledge of different reading strategies (e.g., rapid view of the text, analytic and selective reading, rereading) and their ability to use the appropriate strategy, depending on reading purpose and on materials. Children were credited with 1 point for each correct answer; therefore, possible scores range from 0 to 10 for both subtests. Cronbach's alphas for these scales were acceptable (αs = .74 and .64, respectively). Examples of items are given in the Appendix.

Mental state knowledge. Participants were asked to write an ending for two mentalizing stories concerning episodes of daily life. The first story concerned school life and required the child to impute a false belief; the second story concerned free time and required an understanding of different beliefs (see the Appendix). No time limit was set. Written answers were then coded for the frequency of MSW according to a list of mental state terms previously selected in other studies (Hughes et al., 2007). Examples of the coding categories are given in the Appendix. Each reference to a mental state was summed into a total MSW score. Subsequent repetitions of the same mental state term (i.e., ''She thinks and re-thinks . . .'') and comments that were extraneous to the narratives (i.e., ''I don't know''; ''I like this story'') were excluded. Two researchers independently coded 20% of the stories. Interrater agreement, established using Cohen's kappa, was excellent ($\kappa = .86$). We were interested in the raw number of children's references to mental states rather than the proportion. As Ruffman and colleagues argued (Ruffman, Slade, & Crowe, 2002), each mention to an inner state potentially reflects child's own tendency to refer to mental states to explain other's behavior. In order to control for story length, we calculated the total number of words (excluding mental state terms) in the story and created a nonmental state words (NMSW) total score. As in Ruffman et al. (2002) and Peterson and Slaughter (2006), we considered this variable (NMSW) as a measure of children's expressive language.

Verbal ability. The NMSW score from the story task was used to assess children's expressive language. Children's receptive language was measured using the Italian version of the vocabulary subtest of the Primary Mental Abilities (PMA; Thurstone & Thurstone, 1962; Rubini & Rossi, 1982). This test required children to find the synonym of a target word choosing between four alternative words (30 verbal items), and to match a picture to a word read by the researcher choosing between four alternatives (30 picture-matching items). A time limit of 5 min was set. Performance was evaluated in terms of the number of correct answers. Reliability was good for both subscales (Cronbach's alphas = .83 and .93, respectively).

Reading comprehension task. Two parallel stories, one for second and one for fourth graders from the memory and transfer (MT) standardized reading comprehension battery (Cornoldi & Colpo, 1995) were administered. Research carried out on these age groups has demonstrated good test validity and reliability (for a review, see Cornoldi & Colpo, 1995; Palladino, De Beni, & Cornoldi, 1995). Participants were required to read silently and answer a series of multiple-choice questions, the final score being the number of correct answers. All questions probed inferential information, (e.g., main idea, cause–effect relation, implied meanings) rather than literal information. The procedure allowed participants to have the passages in front of them in order to consult them during the answering phase. No time limit was imposed. Possible scores range from 0 to 10 points. Cronbach's alpha was acceptable, $\alpha = .68$.

Results

We begin by presenting data reduction and descriptive statistics for reading comprehension, receptive and expressive language, metaknowledge about reading and MSW. Before describing associations between variables, we summarize findings concerning age and gender effects. For all correlation analyses between metaknowledge about reading and MSW, we controlled for both receptive and expressive language as well as for reading comprehension.

Data Reduction

The two subtests of metaknowledge, goals of reading and reading strategies, were correlated with each other both in second graders, $r_{\mathrm{II}}(74) = .34$, $p < .01$, and fourth graders, $r_{\mathrm{IV}}(87) = .31$, $p < .01$. Therefore, the scores on the two subtests were summed into an overall index of metaknowledge about reading. Reliability for this aggregate was satisfactory, $\alpha = .66$. Regarding the receptive vocabulary measure, children's performance on the verbal and figural items of the PMA were highly associated both in second graders, $r_{\mathrm{II}}(86) = .46$, $p < .001$, and fourth graders, $r_{\mathrm{IV}}(95) = .57$, $p < .001$. Therefore, an aggregate index of children's receptive vocabulary was created by summing scores on the verbal and figural items. Cronbach's alpha for this overall score was .93, indicating excellent internal consistency. Finally, the frequencies of children's MSW used in the two stories were significantly related to each other, $r_{\mathrm{II}}(86) = .49$, $p < .001$; $r_{\mathrm{IV}}(95) = .33$, $p < .001$. Similarly, the score of NMSW in Story 1 significantly related to the score of NMSW in Story 2, $r_{\mathrm{II}}(86) = .75$, $p < .001$; $r_{\mathrm{IV}}(95) = .79$, $p < .001$. Therefore, aggregate MSW and NMSW scores were created by summing children's scores on the two stories.

Descriptive Statistics

Table 1 presents the means and standard deviations for all raw scores. Next, we conducted a series

Table 1

Descriptive Statistics for Study 1 Measures in Second and Fourth Graders

	M (*SD*) for second graders	*M* (*SD*) for fourth graders	*t*	*df*
Reading comprehension	6.68 (2.49)	7.14 (2.12)	1.27	163
Receptive vocabulary	32.03 (8.66)	41.22 (5.87)	8.51**	183
Nonmental state words (NMSW)	86.41 (55.82)	245.76 (99.02)	11.24**	182
Metaknowledge about reading	9.63 (3.27)	13.25 (2.48)	8.07**	163
Mental state words (MSW)	1.35 (1.65)	3.99 (2.66)	8.00**	183

$^{**}p < .01$ (all significance tests are two-tailed).

of *t* tests to examine effects of age (second versus fourth grade) and gender on children's receptive vocabulary, NMSW, metaknowledge about reading, and MSW. Analyses showed significant main effects of age on all these variables, with fourth graders reporting higher scores than second graders (see Table 1). No significant gender differences were found for three of four variables. The only exception was MSW, $t(181) = 2.59$, for which girls ($M = 3.19$, $SD = 2.7$) showed higher levels than boys ($M = 2.2$, $SD = 2.3$). However, the effect size, $\eta^2 = .03$, suggests that this difference should be reexamined with a larger sample size.

Relation Between Metaknowledge About Reading, Language, and MSW

Table 2 presents the zero-order correlations between age, reading comprehension, receptive vocabulary, NMSW, MSW, and metaknowledge about reading for the sample as a whole and for the second and fourth graders considered separately. As expected, for the total sample, children's age, reading comprehension, and language (both receptive and NMSW) were significantly interrelated and were associated with MSW and metaknowledge about reading. Children who were older and who reported higher scores on language and reading comprehension had a more mature metaknowledge about reading and produced higher frequencies of MSW. Interestingly, the association between metaknowledge about reading and the frequency of MSW was significant even after controlling for receptive vocabulary and NMSW, $r(156) = .19$, $p < .05$; reading comprehension, $r(151) = .39$, $p < .01$; and age, $r(148) = .29$, $p < .01$. When language, reading comprehension and age were controlled together, the correlation between metaknowledge about reading and MSW still approached significance, $r(135) = .16$, $p < .06$. In other words, children who often referred to the characters' mental states were also more likely to

Table 2

Study 1 Zero-Order Correlations Between Study 1 Measures in the Sample Overall and in Second and Fourth Graders Considered Separately

Subscale		Age	RC	RVoc	NMSW	MSW	MKR
Age	All	–	.19*	.54**	.68**	.50**	.45**
	2nd graders		.15	.02	.12	.02	.15
	4th graders		.08	.02	.01	.14	.18
Reading comprehension (RC)	All		–	.52**	.25*	.25*	.43**
	2nd graders			.65**	.16	.13	.38**
	4th graders			.43**	.30**	.30**	.41**
Receptive vocabulary (RVoc)	All			–	.53**	.47**	.56**
	2nd graders				.32**	.33**	.43**
	4th graders				.22*	.25*	.31*
Nonmental state words (NMSW)	All				–	.69**	.45**
	2nd graders					.63**	.22*
	4th graders					.53**	.26*
Mental state words (MSW)	All					–	.47**
	2nd graders						.32**
	4th graders						.28**
Metaknowledge about reading (MKR)							–

$^{*}p < .05$. $^{**}p < .01$ (all significance tests are two-tailed).

obtain higher scores on the Metacomprehension Questionnaire and vice versa. Furthermore, the pattern of correlations shown in Table 2 was very similar for second and fourth graders when examined separately. The main exception was the lack of associations with age. At first glance, this result is surprising; however, note that within each sample the variability in age was restricted.

In order to be sure that children's MSW and NMSW can be considered as two distinct variables, with distinct implications, we also computed a partial correlation between NMSW and metaknowledge about reading, removing the influence of MSW. Interestingly, this partial correlation was not significant, supporting our view that the NMSW variable is a measure of linguistic competence and MSW variable in itself is indeed the crucial index of children's mentalizing abilities.

To summarize, our analyses showed a significant and positive association between second and fourth graders' use of MSW and their metaknowledge about reading, independent of language and reading comprehension.

Discussion

The main aim of Study 1 was to investigate the relation between two domains of children's knowledge about the mind, metaknowledge of reading and mental state knowledge. The critical finding of this study was the significant association between children's metaknowledge about reading and their mental state knowledge as indexed by the tendency to use MSW when completing two stories (specifically chosen so as to include a mentalistic prompt). That is, children who understand that there may be different goals when reading and who can recognize the correct strategies according to the specific task presented, tended to use MSW more frequently in a written story task. Interestingly, this result was found in second and fourth graders, indicating that although performance on tasks assessing both constructs develops significantly between second and fourth grade, as reported in the literature (Carpendale & Chandler, 1996) and in our own data, the link between the two does not seem to change. Thus, our data support the view of a consistent relation between metaknowledge about reading and MSW that is only partially due to VA and reading comprehension. Results from Study 1 fit well with the theoretical framework proposed by Kuhn (1999, 2000) in which mental state knowledge and metacognition are viewed as distinct skills as part of a general ability of "meta-knowing" in which children reflect on their own and other's minds and cognitive processes. They are also consistent with those of Lockl and Schneider (2007), who demonstrated that 4- and 5-year-old children's performance on false belief tasks was associated with their knowledge about memory processes when they were nearly 6.

Although informative and novel, Study 1 had a number of weakness concerning the measure of mental state knowledge. In Study 2, we differentiated between children's cognitive versus emotion mental state knowledge. Although understanding the mental states of others encompasses an understanding of beliefs and knowledge as well as multiple internal states such as desires and emotions (Dunn, 1995; Wellman & Liu, 2004), a growing body of research shows a lack of relation between concurrent cognitive and emotion understanding when language and background variables are taken into account (Cassidy et al., 2003; Cutting & Dunn, 1999; Hughes et al., 2007). To date, however, no study has directly compared concurrent and longitudinal associations between cognitive versus emotion mental state knowledge and metacognitive knowledge. Thus, the first aim of this study was to examine whether cognitive and emotion understanding showed a similar pattern of associations with metacognitive knowledge. Consistent with those studies presented above demonstrating an independence between children's cognitive and emotion understanding and the close link between children's false belief and metacognition, we hypothesized that metacognitive knowledge would be more strongly related to their knowledge about cognitive than emotional mental states.

More important, although Study 1 demonstrated a clear relation between children's mental state knowledge and metaknowledge of reading, its cross-sectional nature meant that we could not be confident that the influence of mental state knowledge on metacognitive knowledge was indeed a causal one. Study 2 was therefore, designed to expand on the findings of Study 1 using a longitudinal design to examine the direction of influence between these two constructs. To this end, we measured mental state knowledge and metaknowledge about reading at two time points. This methodological design allowed us to directly compare two models: the first one (Model A) in which early mental state knowledge predicted later metaknowledge, and the second one (Model B) in which early metaknowledge predicted later mental state knowledge. In order to control for the mediating role of VAs, we covaried children's receptive and expressive language. We did not include children's reading comprehension as a covariate as we found it in Study 1 to play a minor role in the association between mental state knowledge and metaknowledge of reading. Furthermore, in Study 2 the measure of metaknowledge was expanded by pairing a second standardized measure with the questionnaire of metacomprehension to improve the reliability of the measure and to include measures of metaknowledge related to a broader range of school tasks.

Study 2

Method

Participants

In total, 71 children (37 boys and 34 girls) took part in this longitudinal study. None of these children had been previously involved in Study 1. Children were all White and Italian speaking and were recruited from two primary schools with socially mixed catchment areas in the North of Italy. All children had written parental consent to participate and were tested at two time points, separated by a testing interval of approximately 11 months. Children's mean age was 10 years (*SD* = 0.54, range = 9.08–11.16 years) at Time 1 and 10 years 11 months (*SD* = 0.49, range = 10–12 years) at Time 2. Children with learning disabilities were excluded. All 71 children took part in the study at both time points.

Materials and Procedure

At each time point, children participated in two group sessions and in an individual session. The order of the tasks within each session was fixed. In the first group session children were administered the metaknowledge about reading task followed by the story task. The second group session consisted of the Strange Stories task, the Questionnaire of Metacognition and the vocabulary task. In the individual session children's emotion understanding on the Test of Emotion Comprehension (TEC) was evaluated.

Metaknowledge about reading and school tasks. At each time point, children's metaknowledge was assessed via two age-appropriate subtests from the Metacomprehension Questionnaire, as in Study 1 (Pazzaglia et al., 1994), and the Questionnaire of Metacognition (Friso, Palladino, & Cornoldi, 2006).

In the Metacomprehension Questionnaire children were administered the knowledge of reading strategy subtest (10 items) and the text sensitivity subtest (12 items), as they are the most discriminative aspects of metaknowledge on reading at this age (De Beni & Pazzaglia, 1994; see the Appendix for examples). For details on the reading strategies subtest, see Study 1. The text sensitivity subscale was made up of 12 items evaluating children's ability to classify some features of a text according to an importance order, to indicate literary genders (e.g., fairy tale, article, and drama), to evaluate the relative difficulties of two different texts, and to use the title of a text to anticipate its content. Children were credited with 1 point for each correct answer; therefore, possible scores range from 0 to 12 points. Cronbach's alphas for these scales were acceptable ($\alpha = .69$ for reading strategy and $\alpha = .72$ for text sensitivity).

The Questionnaire of Metacognition has been recently developed within a metacognitive training program for primary school children (Friso et al., 2006). It assesses children's metaknowledge not only in relation to the specific context of reading but also to other school tasks such as studying and problem solving. It consists of 15 items tapping different aspects of metaknowledge: the concept of mind, knowledge of strategies in typical learning situations (such as reading, writing, and counting), and knowledge related to control and monitoring. Children were asked to read all the questions silently and to choose the correct answer from four alternatives. No time limit was imposed. One point was given for each correct answer; thus, possible scores ranged from 0 to 15 points. Internal consistency was acceptable (Cronbach's alpha = .68). Examples of some items are given in the Appendix.

Cognitive mental state knowledge. At both time points, children's cognitive understanding was assessed through the Strange Stories task (Happè, 1994) and the frequency of CSW in the story task of Study 1. For time reasons, we included only one of the two stories used in Study 1. We chose Story 1 as it was found to be the one that elicited the highest frequency of MSW (see Study 1 for details on coding).

The Strange Stories task is an advanced test of ToM that assesses the ability to make inferences about mental states by interpreting nonliteral statements (Happè, Winner, & Brownell, 1998; Maylor, Moulson, Muncer, & Taylor, 2002). We selected a subgroup of six mentalistic stories involving cognitive mental states. They included two double bluffs, two mistakes, and two pretence stories. All stories were followed by two questions; a comprehension question ("Is it true what X says?") and a justification or mental state question ("Why does X say this?''). Children were asked to carefully read the stories for themselves and then write the answers; no time limit was set. Justification answers were scored according to their references to mental versus physical states. In line with other studies (Happè et al., 1998; Sullivan & Ruffman, 2004), answers were scored according to a 0–2 scale, giving 0 for an incorrect or did not know answer, 1 for partial or implicit correct answers, and 2 for full and explicitly correct answers. Internal consistency for this measure was good (Cronbach's alpha = .81). An example of the scoring criteria is given in the Appendix. Interrater agreement with the second author, established using Cohen's kappa, was good (κ = .82).

Emotion mental state knowledge. At both time points, emotion understanding was evaluated via the frequency of ESW produced in the story task (see Study 1) and the Italian version of the TEC (Albanese & Molina, 2008).

The TEC is an extensive measure of emotion comprehension in children between 3 and 11 years of age (Pons, Harris, & de Rosnay, 2004). The TEC gives an overall score of emotion understanding evaluating nine components with a hierarchical organization (for more details on these nine components, see the Appendix). It consisted of an A4 picture book with a simple scenario on each page. While showing each picture scenario, the experimenter read the accompanying story with a neutral tone of voice. After hearing the story, children were asked to attribute an emotion to the main character (whose face was blank) by pointing to one of four alternative emotional outcomes depicted below the scenario (two negative and two nonnegative emotions). One point was given for each component correctly answered. Scores ranged between 0 and 9. Cronbach's alpha for this scale was .84.

Verbal ability. As in Study 1, children's receptive language was assessed via the Vocabulary subtest of the PMA (Thurstone & Thurstone, 1962) and children's expressive language was assessed via the total NMSW produced in the story task. The same tasks were administered at both time points. In this study, because of time constraints, we only administered the verbal scale of the Vocabulary subtest. A time limit of 4 min was set. The reliability was good, Cronbach's alpha = .87.

Results

We first present preliminary analyses concerning (a) data reduction and descriptives, (b) developmental changes and stability for all variables, (c) associations between cognitive (Strange Stories, CSW) and emotion mental state knowledge (TEC and ESW), and (d) associations between language and both mental state knowledge and metaknowledge about reading. In the Results section, we address the two focal issues of this study, the specificity and the direction of the relation between children's mental state knowledge and metaknowledge about reading, presenting correlations and hierarchical regression analyses within and between time points. As in Study 1, we give zero-order (in Table 4) as well as partial correlations (in the text) controlling for both receptive and expressive language.

Data Reduction

Children's scores on the two areas of metaknowledge about reading from the Metacomprehension Questionnaire, goals of reading and reading strategies, and on the Questionnaire of Metacognition were significantly associated both at Time 1, mean $r(69) = .46$, rs ranged from .41 to .50, $p < .01$, and at Time 2, mean $r(69) = .33$, rs ranged from .28 to .37, $p < .01$. Therefore, we calculated children's percentages of success in each questionnaire and then averaged these percentages to create an overall measure of metaknowledge. This aggregate showed good internal consistency, Cronbach's alpha = .71.

Descriptive Statistics and Developmental Changes and Continuities

Table 3 shows the descriptive statistics for the study measures at both time points. In order to examine changes across time, we conducted seven paired-sample t tests, applying Holm's correction to maintain familywise error at 0.05. There were significant age-related increases in receptive vocabulary, NMSW and the TEC. No time effect was found for children's references to emotions (ESW), Strange Stories, and metaknowledge about reading, and a significant difference, but in the opposite direction, was found for children's reference to cognitive states (CSW).

Table 4 presents the zero-order concurrent and longitudinal correlations between, language, cognitive mental state knowledge, emotion mental state knowledge, and metaknowledge about reading. The correlations between Time 1 and Time 2 indicate good stability over time for receptive and expressive language, cognitive mental state knowledge (both Strange Stories and CSW), TEC, and metaknowledge about reading, mean $r = .36$, $p < .01$.

Associations Between Cognitive and Emotion Mental State Knowledge

Overall, the pattern of correlations in Table 4 indicated that 9 of 16 correlations between cognitive and emotion mental state knowledge were not significant. When we controlled for receptive and expressive language, however, there were no significant associations, with the exception of the correlation between Time 1 Strange Stories and Time 2 ESW, $r(67) = .28$, $p < .05$. These results support the view of an independence between cognitive and emotion mental state knowledge. Therefore, they will be considered separately in the following analyses. Here, it is also important to note that the associations between CSW and the Strange Stories measures were significant, highlighting the validity of our measures. In contrast, no associations were found between children's frequency of ESW in the story task and their ability to infer others' emotions on the TEC.

Table 3

Descriptive Statistics of Key Measures of Study 2 at Time 1 and Time 2

	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	*t*	*df*
Receptive vocabulary	25.96 (5.25)	28.38 (2.82)	4.67**	70
Nonmental state words (NMSW)	68.80 (34.51)	88.54 (22.49)	4.63**	70
Strange stories	4.49 (1.71)	4.92 (1.73)	2.03	70
Cognitive state words (CSW)	2.45 (2.02)	1.77 (1.38)	−2.69*	70
Test of emotion comprehension (TEC)	7.23 (1.13)	7.78 (1.06)	2.35*	70
Emotion state words (ESW)	1.67 (1.45)	1.69 (1.45)	.06	70
Metaknowledge about reading	0.67 (0.11)	0.70 (0.11)	−1.97	70

*p < .05. **p < .01 (all significance tests are two-tailed).

Associations Between Language, Mental State Knowledge, and Metacognitive Knowledge

Table 4 also shows the correlations between language, mental state knowledge and metaknowledge about reading. Individual differences in children's receptive vocabulary and NMSW were not significantly associated to one another within time points and showed a different pattern of correlations with the four mental state knowledge measures. At both time points, children's NMSW was associated with the frequency of cognitive, $rs(69) \geq .22$, $ps < .05$, and

Table 4

Zero-Order Correlations Between Key Measures of Study 2 Within and Across Time

		Time 1							Time 2						
		RVoc	NMSW	SS	CSW	TEC	ESW	MKR	RVoc	NMSW	SS	CSW	TEC	ESW	MKR
Time 1	Receptive vocabulary (RVoc)	–	.04	.20	.29*	.17	−.09	.36**	.56**	.25*	.15	.21	.36**	.14	.30**
	Nonmental state words (NMSW)		–	.22*	.45**	.08	.37**	.22*	.22*	.22*	.16	.22*	−.01	.23*	.22*
	Strange stories (SS)			–	.39**	.23*	.25*	.29*	.12	.25*	.48**	.24*	.31*	.28*	.31**
	Cognitive state words (CSW)				–	.17	.07	.39**	.19	.26**	.24*	.26**	.26*	.33**	.42**
	Test of Emotion Comprehension (TEC)					–	−.16	.12	.12	−.06	.27*	.01	.39**	−.03	−.04
	Emotion state words (ESW)						–	.10	−.13	−.01	.14	.16	−.19	.12	.05
	Metaknowledge about reading (MKR)							–	.32**	.16	.20	.24*	.35**	.04	.50**
Time 2	Receptive vocabulary (RVoc)								–	.14	.12	.13	.40**	.05	.45**
	Nonmental state words (NMSW)									–	.06	.38**	.03	.39**	.31**
	Strange stories (SS)										–	.33**	.18	.12	.24*
	Cognitive state words (CSW)											–	.20	.13	.33**
	Test of Emotion Comprehension (TEC)												–	.09	.40**
	Emotion state words (ESW)													–	.26*
	Metaknoweldge about reading (MKR)														–

*p < .05. **p < .01 (all significance tests are two-tailed).

emotion words, *rs*(69) ≥ .23, *ps* < .05, as well as metaknowledge about reading, *rs*(69) ≥ .22, *ps* < .05. On the other hand receptive vocabulary seemed to play a minor role, correlating with metaknowledge about reading within and across time, *rs*(69) ≥ .30, *ps* < .01, yet only with CSW at Time 1 and the TEC at Time 2. Having found that VA was, at least in part, associated with mental state knowledge and metacognitive knowledge, we also controlled for VA in subsequent analyses.

Concurrent Relations Between Mental State Knowledge and Metaknowledge About Reading

We first checked whether we could replicate the results of Study 1 with a different sample and in slightly older children. As in Study 1, partial correlations controlling for language between children's frequency of MSW and their metaknowledge were significant, $r(67) = .26$, $p < .05$ at Time 1 and $r(67) = .31$, $p < .01$ at Time 2. In addition, the relation between NMSW and metaknowledge controlling for MSW was, as in Study 1, not significant at both time points, *rs*(67) ≤ .14, *ns*.

Study 2 was designed to examine whether children's cognitive and emotion mental state knowledge differentially related to their level of metaknowledge about reading. Table 4 shows a different pattern of correlations depending on the time point considered. At Time 1, the relation between metaknowledge about reading and mental state knowledge seems to be different depending on the type of mental state considered. Only cognitive mental state knowledge (both the Strange Stories and CSW), *rs*(69) ≥ .25, *ps* < .05, and not emotion mental state knowledge, *rs*(69) ≤ .12, *ns* (both the TEC and ESW) was positively associated with children's scores on the Metacognitive Questionnaire even after controlling for individual differences in VAs, $r(67) = .24$, $p < .05$. By contrast, at Time 2, metaknowledge about reading was found to be significantly related to both cognitive and emotion mental state knowledge, even after controlling for VAs, $r_{\text{CSW}}(67) = .23$, $p < .05$; $r_{\text{TEC}}(67) = .29$, $p < .05$.

Longitudinal Relation Between Mental State Knowledge and Metaknowledge

A core aim of Study 2 was to examine the direction of the relation between children's mental state knowledge and their metaknowledge about reading, comparing two opposite models: Model A, in which early mental state knowledge predicts later metaknowledge, and Model B, in which early metaknowledge predicts later mental state knowledge. In order to investigate this issue we performed correlations and hierarchical regression analyses.

Model A. Table 4 shows that at Time 1 both measures of cognitive mental state knowledge were significantly related to Time 2 metaknowledge about reading, *rs*(69) ≥ .31, *ps* < .01. In contrast neither the TEC nor ESW at Time 1 were related to Time 2 metaknowledge of reading. In order to examine the predictive effect of early cognitive mental state knowledge on later metacognitive knowledge, we computed a hierarchical regression analysis (see Table 5). Given the robust correlations between children's scores on CSW and Strange Stories, we aggregated the two measures into an overall index of cognitive mental state knowledge. Time 2 cognitive mental state knowledge aggregate was log transformed in order to reduce any effects of the positive skew that was apparent in the raw data. We entered Time 1 language (receptive vocabulary and NMSW) and metaknowledge about reading at Step 1 and the Time 1 cognitive mental state knowledge composite at Step 2 to establish whether it made a significant independent contribution in

Table 5

Hierarchical Regression Analysis Examining Time 1 Cognitive Mental State Knowledge as a Predictor of Metaknowledge About Reading at Time 2 Controlling for Verbal Abilities and Metaknowledge About Reading at Time 1

	Variable	*B*	*SE B*	ß
Step 1	Vocabulary Time 1	.01	.01	.14
	NMSW Time 1	.01	.01	.12
	Metaknowledge about reading Time 1	.44	.12	.43**
Step 2	Vocabulary Time 1	.01	.01	.09
	NMSW Time 1	.01	.01	.06
	Metaknowledge about reading Time 1	.38	.11	.37**
	Cognitive mental state knowledge Time1	.02	.01	.25*

Note. $R^2 = .28$ for Step 1; $\Delta R^2 = .05$ for Step 2. NMSW = nonmental state words.
**p* < .05. ***p* < .01 (all significance tests are two-tailed).

predicting variance in metaknowledge about reading at Time 2. Table 5 shows that at Step 1, Time 1 metaknowledge about reading was the only significant predictor, accounting for 28% of the variance in Time 2 metacognitive knowledge, $F(3, 67) = 9.43$, $p < .001$. The addition of Time 1 cognitive mental state knowledge significantly improved the amount of variance explained, $\Delta R^2 = .05$, $F(4, 66) = 4.22$, $p < .001$. When the metaknowledge reading items and the metaknowledge of general school activity items were considered separately, the addition of Time 1 cognitive mental state knowledge in the second step still approached significance, $p < .08$ for both regressions.

We also conducted a second analysis to assess whether the effect of Time 1 cognitive mental state knowledge on Time 2 level of metaknowledge about reading was independent of the effect of Time 2 cognitive mental state knowledge. We, therefore, entered children's Time 1 language (receptive vocabulary and NMSW), metaknowledge about reading, and Time 2 cognitive mental state knowledge at Step 1 and Time 1 cognitive mental state knowledge composite at Step 2 of the regression (see Table 6). Step 1 was significant with early metaknowledge about reading and cognitive mental state knowledge being the only significant predictors, $R^2 = .22$, $F(4, 66) = 5.99$, $p < .001$. The addition of Time 1 cognitive mental state knowledge did significantly improve the amount of variance explained in the criterion variable, $\Delta R^2 = .04$, $F(1, 65) = 5.84$, $p < .05$. Interestingly, when we tested the opposite model in order to assess the effect of Time 2 cognitive mental state knowledge on concurrent (Time 2) level of metaknowledge about reading controlling for Time 1 cognitive mental state knowledge, language and metaknowledge about reading, Time 2 cognitive mental state knowledge did not significantly explain a unique percentage of variance in Time 2 metaknowledge about reading, $F(1, 65) = 1.44$, *ns*.

Model B. Table 4 also shows that Time 1 metaknowledge about reading is significantly related with Time 2 CSW and emotion comprehension. In order to better understand these relations, we performed two separate hierarchical regressions with Time 2 CSW and Time 2 emotion comprehension as dependent variables, respectively. In the first regression (Time 1 metaknowledge over Time 2 CSW), we entered Time 1 CSW and language in Step 1 and Time 1 metaknowledge about reading in Step 2. Neither Step 1, $F(3, 66) = 1.92$, *ns*, nor Step 2, $F(1, 65) = 1.03$, *ns*, was significant. In the second regression (Time 1 metaknowledge over Time 2 emotion understanding), we entered Time 1 emotion understanding and language at Step 1 and Time 1 metaknowledge about reading at Step 2. Results showed that Step 1 explained a significant 19% of the variance, $F(3, 66) = 6.28$, $p < .01$, with Time 1 emotion understanding and receptive vocabulary being significant predictors. However, Time 1 metaknowledge about reading did not significantly improve the amount of variance explained in the criterion variable, $\Delta R^2 = .04$, $F(1, 68) = 5.80$, *ns*.

In sum, the results of previous analyses provide support for Model A in which early mental state knowledge predicts later metaknowledge. The main finding of this second study was that there was a predictive relation between children's early cognitive (but not emotion) mental state knowledge and their later metaknowledge, even when continuities in metaknowledge and VAs were taken into account.

Table 6

Hierarchical Regression Analysis Examining Time 1 Cognitive Mental State Knowledge as a Predictor of Concurrent Metaknowledge About Reading at Time 2 Controlling for Verbal Abilities and Metaknowledge About Reading at Time 1 and Cognitive Mental State Knowledge at Time 2

	Variable	*B*	*SE B*	ß
Step 1	Vocabulary Time 1	.01	.00	.17
	NMSW Time 1	.00	.00	.14
	Metaknowledge about reading Time 1	.29	.12	.29*
	Cognitive mental state knowledge Time 2	.37	.18	.23*
Step 2	Vocabulary Time 1	.01	.00	.12
	NMSW Time 1	.00	.00	.05
	Metaknowledge about reading Time 1	.27	.11	.27*
	Cognitive mental state knowledge Time 2	.22	.19	.14
	Cognitive mental state knowledge Time 1	.02	.01	.26*

Note. $R^2 = .22$ for Step 1; $\Delta R^2 = .04$ for Step 2. NMSW = nonmental state words.
*$p < .05$ (all significance tests are two-tailed).

Discussion

The purpose of this second study was twofold. Our first goal was to examine the specificity of the link between mental state knowledge and metaknowledge about reading by distinguishing two components of children's mental state knowledge: cognitive and emotion mental state knowledge. Our second goal was to investigate the direction of the link between children's mental state knowledge and their metaknowledge about reading adopting a longitudinal design. The strengths of Study 2 are that, first, it allowed us to look at the predictive effect of mental state knowledge on later metaknowledge about reading. Second, it adopted a finer grain analysis of mental state knowledge. Third, it integrated the analysis of the use of MSW with more classic measures of mental state knowledge. Finally, it included additional measures of metaknowledge about reading.

Before discussing the main findings of Study 2 there are preliminary results that warrant discussion. First, it is important to note that, at least in part, findings of Study 2 replicated, in slightly older children, the results of Study 1. That is, children who referred more to mental states in the story task also reported higher scores on metaknowledge about reading tasks, independently of VAs. In other words, as expected on the basis of our first study, there was a significant amount of shared variance between children's understanding of how their own minds work in relation to school tasks and their tendency to interpret others' behavior in terms of mental states. Second, we observed stability across time in children's mental state knowledge and metacognitive knowledge, even when language was taken into account. This finding echoes those of similar studies that demonstrated stability of individual differences in children's emotion (Pons & Harris, 2005) and false belief understanding (Hughes & Dunn, 1998) and in their beliefs about themselves as readers (Bouffard, 1998). Third, there were few significant associations between mental state knowledge and language. Although this finding needs further replication, we suggest that it could reflect the age of our sample. Indeed, previous research shows that individual differences in MSW and language are strongly associated in preschoolers (Bretherton & Beeghly, 1982; Hughes et al., 2006) but not in school-aged children and adolescents (Meins et al., 2006; O'Connor & Hirsch, 1999). Therefore, it is likely that the association between mental state knowledge and language shows a developmental trajectory.

The first main aim of this study was to investigate the specific link between mental state knowledge and metaknowledge about reading by examining children's cognitive versus emotion mental state knowledge. Motivation for this comparative analysis was confirmed with the finding that, with the exception of the relations between Strange Stories and emotion understanding at Time 1, there were no significant correlations between cognitive and emotion mental state knowledge at both time points. This result fits with studies on younger children (see the Introduction) and with neuroimaging studies reporting distinct anatomical correlates for understanding of thoughts versus feelings (see Saxe, Carey, & Kanwisher, 2004).

Regarding the specificity of the relation, the present study showed a differential relation between cognitive mental state knowledge and children's metaknowledge about reading. That is, as we predicted, the predictive effect of early mental state knowledge on later metaknowledge about reading was found only for cognitive and not for emotion understanding. We argue that the close and specific connection between children's understanding of cognitions and metaknowledge reflects the conceptual link between the two domains. As we have said in the Introduction, children's knowledge of how their minds work when reading presupposes a familiarity with terms indicating cognitive activities and also requires the ability to infer and to reflect on cognitions. On the other hand, children's metaknowledge about reading does not necessarily require a knowledge of emotional states, which seems to be more relevant for social aspects of school readiness (Denham, 2006).

Besides discussing the specificity of the relation, we believe that the lack of association between the two measures of children's emotion mental state knowledge, TEC and ESW, deserves further attention. From a methodological perspective, this result was unexpected and questioned the reliability of the emotions measures. However, it is important to note that children's scores on the TEC were stable across time and consistent with the age-appropriate findings reported in Albanese and Molina (2008) and Pons et al. (2004). Therefore, it is also possible that the lack of association between the TEC and ESW reflects a theoretical issue rather than a methodological concern. It could be explained by considering the difference between the TEC and the story task. Whereas the TEC assesses the level of possession of an ability, (i.e., competence), the production of emotion state terms evaluates children's spontaneous tendency to *use* an ability in interpreting social behavior. An intriguing possibility is that in school-aged children the tendency to focus on internal states can be independent from their capacity to represent those mental states. In line with this interpretation, Meins and colleagues recently argued that "children's proportional use of internal-state language in noninteractional contexts can be thought of more as a measure of children's *spontaneous use* of their ToM capacities than of those capacities themselves" (Meins et al., 2006, p. 193).

Although we find in line with our prediction that emotion mental state knowledge was not predictive of children's metaknowledge of reading, it is worth considering, however, whether other aspects of emotion development such as emotion regulation might have related. Emotion regulation is described as the capacity to modify, regulate or evaluate one's emotional reactions and is influenced by children's developing ToM as well as emotion understanding (see Thompson, 1994; Thompson, Lewis, & Calkins, 2008, for reviews). Furthermore, in older children, development in inhibitory control fosters increased emotion regulation through emotional self-management (Thompson et al., 2008). Since the regulation of emotion requires the development, not only of ToM but also awareness of the self-management of emotion, it is plausible that this skill could relate to ongoing developments in metacognition. This is a future research direction which requires further attention. It is also important to note that, although at Time 1 metaknowledge about reading was linked with cognitive mental state knowledge and not with emotion mental state knowledge, at Time 2 metacognition was equally related with these two areas of mental state knowledge. Therefore, given the predictive role of cognitive mental state knowledge on metacognition, correlates of metaknowledge about reading may change with time. Despite the lack of available data, it is possible that the link between mental state knowledge and metacognition is sensitive to development (Kuhn, 2000). In other words, it could be that being able to represent a cognitive mental state and having a concept of that state is fundamental for children's emerging understanding of how the mind works with respect to specific tasks. Later in development, when mental state knowledge and metacognitive knowledge become more sophisticated (e.g. Flavell, 2000; Perner, 2005), the exchange between the two could become less specific.

The second core aim of this study was to investigate the direction of the relation between children's mental state knowledge and their metaknowledge about reading. Our main results indicated that children's cognitive mental state knowledge, as assessed via the Strange Stories task and the frequency of cognitive state terms in a story task (but not their emotion mental state knowledge), significantly predicted later metaknowledge about reading even when we controlled for children's earlier metacognitive knowledge, receptive and expressive language and Time 2 cognitive mental state knowledge. In contrast, early metaknowledge about reading did not predict children's later mental state knowledge. Moreover, children's concurrent cognitive mental state knowledge did not explain metaknowledge when earlier cognitive mental state knowledge was controlled for. Taken together these results provide strong support for the causal relation between cognitive understanding and the development of metaknowledge relevant to school tasks. Therefore, early mental state knowledge can be considered a precursor of subsequent metaknowledge and can be seen as a step in a series of change in children's understanding of knowledge. In addition, although the questionnaire tasks we used in the present studies were selected to specifically focus on metaknowledge of reading, certain questions in the second Metacognitive Questionnaire did address other aspects of metacognitive understanding such as "remembering" and "writing." Preliminary results looking at these more general items suggest that our findings do extend to metacognitive awareness within other school tasks more generally, although this needs further research.

Our findings fit nicely with the argument that children's mental state knowledge is a prerequisite for the development of their metacognition (Bartsch & Estes, 1996; Wellman, 1985). In other words it seems that children's capacities to mentalize about reality and to infer cognitive mental states underlying social actions helps them understand how their own mind works in relation to metaknowledge of reading. On this view, children have to acquire an understanding of mental states before they can reflect on their own or on other people's cognitive activity (Perner, 2000), and a representation of the concept of mental state is necessary to develop the ability to think and regulate one's own mental activity in relation to a specific cognitive task (Bartsch & Estes, 1996). Our results are also consistent with prior studies that have showed significant relations between children's mental state knowledge and indirect measures of metaknowledge such as awareness of the interpretive character of the knowing processes (Carpendale &

Chandler, 1996), metacognitive vocabulary (Antonietti et al., 2006), and understanding that the teaching process depends on a knowledge difference between teachers and pupils (Astington & Pelletier, 1996; Ziv & Frye, 2004). Finally, the results of the present study provide support for Lockl and Schneider (2007) who argued that "theory of mind is a general cognitive tool able to influence cognitive processes in a variety of domains" (p. 164).

Conclusion

To summarize, this study examined school-aged children's capacity to reason about mental states by examining the concurrent and longitudinal relation between children's mental state knowledge and metaknowledge about reading. Generally speaking, results indicate that (a) there is a strong link between school-aged children's mental state knowledge and metaknowledge about reading, (b) the link between mental state knowledge and metacognition is causal and goes from earlier mental state knowledge to later metaknowledge about reading, (c) the link between earlier mental state knowledge and later metaknowledge about reading is independent of language and reading comprehension, and (d) the link between earlier mental state knowledge and later metaknowledge about reading is specific to cognitive mental state knowledge. These findings are original yet consistent with previous studies and postulated models. This is an important contribution to the existing literature, since both mental state knowledge and metacognitive knowledge are crucial for children's cognitive and social development. Finally, we believe these findings have implications for intervention and could therefore contribute to promoting efforts to connect metacognition and knowledge of mental states in an educational perspective (Gonida & Kiosseoglou, 2006).

References

Albanese, O., & Molina, P. (2008). *Lo sviluppo della comprensione delle emozioni e la sua valutazione. La standardizzazione italiana del Test di Comprensione delle Emozioni.* [The development of emotion understanding and its evaluation. The Italian Standardization of Test of Emotion Comprehension]. Milano, Italy: Unicopli.

Antonietti, A., Liverta-Sempio, O., Marchetti, A., & Astington, J. (2006). Mental language and understanding of epistemic and emotion mental states. In A. Antonietti, O. Liverta-Sempio, & A. Marchetti (Eds.), *Theory of mind and language in developmental context* (pp. 1–30). New York: Springer.

Astington, J. W., & Pelletier, J. (1996). The language of mind: Its role in teaching and learning. In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 593–620). Malden, MA: Blackwell.

Baker, L. (1994). Fostering metacognitive development. In H. Reese (Ed.), *Advances in child development and behaviour* (pp. 201–239). San Diego, CA: Academic Press.

Bartsch, K., & Estes, D. (1996). Individual differences in children's developing theory of mind and implications for metacognition. *Learning and Individual Differences, 8*, 281–304.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Bouffard, T. (1998). A developmental study of the relationship between reading development and the self-system. *European Journal of Psychology of Education, 13*, 61–74.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, 6*, 906–921.

Brown, A. L., Armbruster, B. B., & Baker, L. (1986). The role of metacognition in reading and studying. In J. Orasanu (Ed.), *Reading comprehension from research to practice* (pp. 49–75). Hillsdale, NJ: Erlbaum.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harward University Press.

Camaioni, L., Longobardi, E., & Bellagamba, F. (1998). Evoluzione dei termini di stati mentali nelle storie di fantasia scritte da bambini in età scolare [The development of terms of mental states in fantasy stories written by school-aged children]. *Età Evolutiva, 60*, 20–29.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686–1706.

Carpendale, J. I. M., & Lewis, C. (2006). *How children develop social understanding*. Oxford, UK: Blackwell.

Cassidy, K. W., Werner, R. S., Rourke, M., Zubernis, L. S., & Balaraman, G. (2003). The relationship between psychological understanding and positive social behaviors. *Social Development, 12*, 198–221.

Cornoldi, C., & Colpo, G. (1995). *Nuove prove di lettura MT* [New MT tests of reading]. Firenze, Italy: Organizzazioni Speciali.

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 94–113). Mahwah, NJ: Erlbaum.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70*, 853–865.

Day, J., Cordon, L. A., & Kerwin, M. L. (1989). Informal instruction and the development of cognitive skills. In C. B. McCormick, G. E. Miller, & M. Pressley (Eds.), *Cognitive strategy research: From basic research to educational applications* (pp. 83–103). New York: Springer-Verlag.

De Beni, R., & Pazzaglia, F. (1994). Analisi di alcune abilità metacognitive implicate nella comprensione del testo scritto: Caratteristiche evolutive e relazioni tra competenza metacognitive e abilità di lettura. *Rassegna di Psicologia*, *1*, 77–97.

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, *17*, 57–89.

Denham, S. A., McKinly, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioural predictors of preschool peer ratings. *Child Development*, *61*, 1145–1152.

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition and Emotion*, *9*, 187–201.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development*, *62*, 1352–1366.

Dyer, J. R., Shatz, M., & Wellman, H. M. (2000). Young children's storybooks as a source of mental state information. *Cognitive Development*, *15*, 17–37.

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, *24*, 15–23.

Flavell, J., Green, F. L., & Flavell, E. R. (1998). The mind has a mind of it s own: Developing knowledge about mental uncontrollability. *Cognitive Development*, *13*, 127–138.

Flavell, J., Green, F. L., & Flavell, E. R. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, *1*, 97–112.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.

Fox, R. (1991). Developing awareness of mind reflected in children's narrative writing. *British Journal of Developmental Psychology*, *9*, 281–298.

Friso, G., Palladino, P., & Cornoldi, C. (2006). Questionario "Io e la mia mente" [Metacognitive Questionnaire]. In G. Friso, P. Palladino, & C. Cornoldi (Eds.), *Avviamento alla metacognizione* (pp. 217–221). Trento, Italy: Erikson.

Garner, R. (1987). *Metacognition and Reading Comprehension*. Norwood, NJ: Ablex.

Gonida, E., & Kiosseoglou, G. (2006, July). *The development of metacognition in preschool age: The role of the theory of mind*. Paper presented at the meeting of the EARLI SIG in Metacognition, Cambridge, UK.

Happè, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feeling by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 128–153.

Happè, F., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Development Psychology*, *34*, 358–362.

Hughes, C., & Dunn, J. (1997). Pretend you didn't know: Preschoolers' talk about mental states in pretend play. *Cognitive Development*, *12*, 381–403.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*, 1026–1037.

Hughes, C., Fujisawa, K. K., Ensor, R., Lecce, S., & Marfleet, R. (2006). Cooperation and conversations about the mind: A study of individual differences in 2-year-olds and their siblings. *British Journal of Developmental Psychology*, *24*, 53–72.

Hughes, C., Lecce, S., & Wilson, C. (2007). "Do you know what I want?" Preschoolers talk about desires, thoughts and feelings in their conversations with sibs and friends. *Cognition and Emotion*, *21*, 330–350.

Kamavar, D., & Homer, B. (2000). Internal and external notions of metarepresentation: A developmental perspective. In J. Astington (Eds.), *Minds in the making: Essays in honour of David Olson* (pp. 197–211). Oxford, UK: Blackwell.

Kolić-Vehovec, S., & Bajšanski, I. (2007). Comprehension monitoring and reading comprehension in bilingual students. *Journal of Research in Reading*, *30*, 198–211.

Kuhn, D. (1999). Metacognitive development. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 259–286). Philadelphia: Psychology Press.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 301–326). Hove, UK: Psychology Press.

Lockl, K., & Schneider, W. (2007). Knowledge about the mind: Links between theory of mind and later metamemory. *Child Development*, *78*, 148–167.

Maylor, E. A., Moulson, J. M., Muncer, A. M., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline in old age? *British Journal of Psychology*, *93*, 465–485.

McElwain, N. L., & Volling, B. L. (2002). Relating individual control, social understanding, and gender to child friend interaction: A relationships perspective. *Social Development*, *11*, 362–385.

Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology*, *24*, 181–196.

O'Connor, T. G., & Hirsch, N. (1999). Intra-individual difference and relationship-specificity of mentalizing in early adolescence. *Social Development*, *8*, 256–274.

Orasanu, J. (1986). *Reading comprehension from research to practice*. Hillsdale, NJ: Erlbaum.

Palladino, P., De Beni, R., & Cornoldi, C. (1995). L'evoluzione delle difficoltà specifiche di comprensione del testo: uno studio longitudinale [The development of reading comprehension difficulties: A longitudinal study]. *Archivio di Psicologia, Neurologia e Psichiatria, 56*, 548–564.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 7*, 1239–1252.

Pazzaglia, F., Cornoldi, C., & Tressoldi, P. E. (1993). Learning to read: Evidence on the distinction between decoding and comprehension skills. *European Journal of Psychology of Education, 8*, 247–258.

Pazzaglia, F., De Beni, R., & Cristante, F. (1994). *Prova di metacomprensione* [Metacomprehension task]. Firenze, Italy: Organizzazioni Speciali.

Pelletier, J., & Astington, J. W. (2004). Action, consciousness and theory of mind: Children's ability to coordinate story characters' actions and thoughts. *Early Education and Development, 15*, 5–22.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Perner, J. (2000). About + Belief + Counterfactual. In P. Mitchell & K. L. Riggs (Eds.), *Children's reasoning and the mind* (pp. 367–401). Hove, UK: Psychology Press.

Perner, J. (2005). Theory of mind in 2005: Old friends and old problems and still a case for conceptual development. *Hellenic Journal of Psychology, 2*, 80–92.

Peterson, C. C., & Slaughter, V. P. (2006). Telling the story of theory of mind: Deaf and hearing children's narratives and mental state understanding. *British Journal of Developmental Psychology, 24*, 151–179.

Pons, F., & Harris, P. L. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition & Emotion, 19*, 1158–1174.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127–152.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behaviour & Brain Sciences, 4*, 515–526.

Rubini, V., & Rossi, M. A. (1982). Analysis of the psychometric characteristics of the Primary Mental Abilities Tests (Level 4-sup-6, 1962 revision) applied to a group of Italian subjects. *Bollettino di Psicologia Applicata, 161–164*, 87–105.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*, 734–751.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87–124.

Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning, 48*, 71–106.

Sullivan, S., & Ruffman, T. (2004). Social understanding: How does it fare with advancing years? *British Journal of Psychology, 95*, 1–18.

Thompson, R. A. (1994). Emotional regulation: A theme in search of definition. *Monographs of the Society for Research in Child Development, 59*(Serial No. 240), 25–52.

Thompson, R. A., Lewis, M., & Calkins, S. (2008). Reassessing emotion regulation. *Child Development Perspectives, 2*, 124–131.

Thurstone, L. L., & Thurstone, T. G. (1962). *Primary mental abilities.* Chicago: Science Research.

Wellman, H. M. (1985). The origins of metacognition. In D. L. Forrest-Pressley, G. E. McKinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 1–31). New York: Academic Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541.

Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development, 19*, 457–477.

Appendix

Example items from the Metacomprehension Questionnaire (Pazzaglia et al., 1994)

Subtest of Reading Goals

In your opinion, when you read, the most important thing is:

A. To read loudly and clearly
B. To read exactly and without hesitation every word
C. To understand the content of the text
D. To read as quickly as possible

Subtest of Reading Strategies

If you don't understand a word inside a text, You can:
A. Try to guess
B. Skip or substitute it with an other word
C. Use the rest of the sentence to understand it

Subtest of Text Sensitivity

Once upon a time there was a prince that woke a princess with a kiss. They lived happily together. This text is: ___________ ___________________________________

Example items from the Questionnaire of Metacognition (Friso et al., 2006)

If you need to remember a lesson/passage it is better:
A. To read it several times
B. To underline with a red pen
C. To remember with your words what happened
D. To study with a TV on in order to keep yourself awake

When you write, the most important thing is:
A. Make the others understand you
B. Not make spelling mistakes
C. Use unusual words
D. Write a text as long as possible

Story Task

Story 1: Luca carries out a school test and tells his mum that he has done it well. The next day he goes to school and the teacher tells him that the test is inadequate. You continue:

Story 2: Alice has been invited to Marta's birthday party and she promised she would go. When she tells her mum at home about the party, her mum reminds her that her cousin was also expecting her at another party. You continue:

Coding Categories for Mental State Words (including verbs, nouns, adverbs and adjectives) and Examples from Children's Transcripts

Cognitive words: ''She forgot to study''; ''The teacher understood that it was not true.''

Emotion words for positive and negative sensations, emotions, feelings, and emotional expressions: ''Mum was happy for the good mark and gave him a present''; ''His heart was full of satisfaction.''

Desire words for feelings of intention, desire, and possibility: ''I wish that you had informed me first''; ''He decided to continue.''

Morality words for moral feelings, emotions, opinion: ''He felt a little bit guilty''; ''He regretted doing this.''

Components of the Test of Emotion Comprehension (Pons et al., 2004)

I. Recognition: the ability to recognize and name emotions on the basis of expressive cues.
II. External Cause: the ability to understand how external causes affect the emotions of other children.
III. Desire: the ability to understand that emotional reactions of other people depend on their desires.
IV. Belief: the ability to understand that person's beliefs determine his emotions.
V. Reminder: the ability to understand that elements of a situation can reactivate past emotions.
VI. Regulation: the ability to understand strategies for emotional control.
VII. Hiding: the ability to understand the discrepancy between the expression of emotion and felt emotion.
VIII. Mixed: the ability to understand that a person may have multiple emotional responses to a given situation.
IX. Morality: the ability to understand that negative feelings ensue from a morally reprehensible action.

Example items from the Strange Stories Task (Happè, 1994)

Katie and Emma are playing in the house. Emma picks up a banana from the fruit bowl and holds it up to her ear. She says to Katie, ''Look! This banana is a telephone.''

Mental State or Justification Question: Why does Emma say this?

Examples responses scored 2: ''She's just pretending.''

Examples responses scored 1: ''Because the banana looks like a telephone.''

Child Development, November/December 2010, Volume 81, Number 6, Pages 1894–1905

Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect

Chiara Turati
Università degli Studi di Milano-Bicocca

Elisa Di Giorgio, Lara Bardi, and Francesca Simion
Università degli Studi di Padova

Holistic face processing was investigated in newborns, 3-month-old infants, and adults through a modified version of the composite face paradigm and the recording of eye movements. After familiarization to the top portion of a face, participants ($N = 70$) were shown 2 aligned or misaligned faces, 1 of which comprised the familiar top part. In the aligned condition, no visual preference was found at any group age. In the misaligned condition, 3-month-olds preferred the face stimulus with the familiar top part, adults preferred the face stimulus with the novel one, and newborns did not manifest any visual preference. Results revealed that both infants' and adults' eye movements may be affected by holistic face information and demonstrated holistic face processing in 3-month-olds.

The human face provides us with critical information about the age, gender, emotional state, intention, and identity of others, thus being one of the most significant visual objects a person encounters. Our ability to recognize faces at a glance requires differentiating subtle information that distinguishes each individual and concerns the shape of facial features (*featural* information), the spacing among features (*second-order configural* information), and the overall structure of the face, where face parts and their relations are glued together in an undifferentiated unit (*holistic* information; Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993). It is generally accepted that the exceptionally proficient face recognition abilities seen in adults are mainly derived from the rapid and efficient encoding of configural information, which includes sensitivity to first-order relations that specifies the stimulus as a face, holistic processing that interconnects facial features in a single unit, and sensitivity to second-order relations that defines differences among individuals in the spacing of facial features (Maurer et al., 2002). The current article focused on holistic processing, investigating through the testing of the composite face illusion newborns' and 3-month-old infants' sensitivity to holistic face information.

Although experience-based developmental processes progressively refine early configural face recognition abilities, available evidence demonstrates that configural processing emerges during the first months of life rather than undergoing a long developmental trajectory. At birth, sensitivity to *first-order* configural information is demonstrated by newborns' preference for faces as compared to equally complex nonface visual stimuli (Johnson & Morton, 1991; Valenza, Simion, Macchi Cassia, & Umiltà, 1996). By the age of 4 months, infants can discriminate faces on the basis of *second-order* configural changes in the spatial relations between facial features (Bhatt, Bertin, Hayden, & Reed, 2005; Deruelle & de Schonen, 1998; Hayden, Bhatt, Reed, Corbly, & Jospeh, 2007; Thompson, Madrid, Westbrook, & Johnston, 2001). Moreover, newborns' and 4-month-old infants' ability to recognize faces is deteriorated by face inversion, suggesting a reliance on configural information in processing upright faces (Turati, Macchi Cassia, Simion, & Leo, 2006; Turati, Sangrigoli, Ruel, & de Schonen, 2004).

Research was supported by grants from Ministero dell'Università (2007XFM93B_004) and University of Padua (2007-CPDA 075245). We thank Dr. Dalla Barba and the nursing staff at the Pediatric Clinic for their collaboration. Many thanks also to Dr. Bruno Rossion and Dr. Adelaide de Heering for sharing the composite face stimuli, and Jennifer Wagner for English editing.

Correspondence concerning this article should be addressed to Chiara Turati, Dipartimento di Psicologia, Università degli Studi di Padova Milano-Bicocca, via dell'Innovazione 10, 20126 Milano, Italy. Electronic mail may be sent to chiara.turati@unimib.it.

© 2010 The Authors
Child Development © 2010 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8106-0020

Finally, 6-month-olds and newborns can discriminate upright Thatcher faces (eyes and mouth reversed on the face) from non-Thatcher faces, even though only relational aspects of the features—second-order configural information—differentiate the two stimuli (Bertin & Bhatt, 2004; Leo & Simion, 2009).

Infants' processing of *holistic* face information has received far less attention with few studies examining infants' ability to process the multiple features of a face as a perceptual unit. The only available evidence showed that 4-month-olds look longer at a "switched" face made up of the internal features of a familiar face and the external features of a different familiar face than at each whole familiar face, indicating that they process the configuration made up of internal and external features (Cashon & Cohen, 2001, 2003). Nevertheless, such studies focused exclusively on the correlations between outer and inner face features. Yet, the inner part of the face per se contains sufficient cues to support 4-month-old infants' face recognition. Newborns use information about internal facial features in making preferences based on attractiveness (Slater et al., 2000) and are able to recognize a face based solely on its inner face part (Turati et al., 2006). Thirty-five- to 40-day-old infants are able to discriminate their mother from a stranger even when the whole line between face and hair is masked by a scarf, thus relying mostly on the inner portion of the face (Bartrip, Morton, & de Schonen, 2001). Following familiarization to unfamiliar faces, similar results have been obtained at 1, 3, and 6 months of age (de Haan, Johnson, Maurer, & Perrett, 2001; Pascalis, de Haan, Nelson, & de Schonen, 1998). Nevertheless, holistic processing of internal features per se, rather than in relation with outer features, has never been tested in infants.

Also, the "switch" paradigm used with infants highly differs from the paradigms commonly employed to test holistic face processing abilities in older children and adults, specifically the part–whole paradigm and the composite paradigm. In the part–whole paradigm, recognition of a face part is measured when presented in the context of the whole face relative to when presented in isolation, and better performance is typically observed for the whole compared to the part condition (Tanaka & Farah, 1993). Pellicano and Rhodes (2003) provided evidence that 4-year-olds manifest a part–whole effect, indicating that they recognize faces holistically. The composite effect is commonly taken as the most compelling evidence that facial features are integrated into a holistic representation of an individual face by human observers and refers to the observation that recognition of the top half of a face is more difficult when the top half is aligned with the bottom half of a different face, creating the impression of a novel face, as compared to when the two halves are misaligned through a lateral shift (Hole, 1994; Young, Hellawell, & Hay, 1987). The difference between the subjects' performance in the aligned and misaligned condition is taken as a measure of the interference produced by holistic processing of the novel composite face on the recognition of its constituent parts. The composite effect has been demonstrated at 6 years of age for both familiar (Carey & Diamond, 1994) and unfamiliar (Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007) faces, and recently in preschool-aged children (de Heering, Houthuys, & Rossion, 2007; Macchi Cassia, Picozzi, Kuefner, Bricolo, & Turati, 2009).

This study addressed holistic face processing in 3-month-old infants (Experiment 1), newborns (Experiment 2), and adults (Experiment 3) through the use of a modified version of the composite face task and the recording of eye movements. The classical composite face paradigm used with adults and preschool- or school-age children was modified in order to be adapted to infants, who are unable to comprehend written or verbal instructions and to provide a manual response. The same paradigm was applied also to adult participants (Experiment 3), because, despite the rich literature on the composite effect in adults, only a single study recorded adults' eye movements during the composite task (de Heering, Rossion, Turati, & Simion, 2008). Thus, Experiment 3 tested whether gaze behavior in adults is affected by the use of holistic face-processing strategies in a composite paradigm analogous to that employed with infants. Newborns, 3-month-old infants, and adults were tested using a visual paired-comparison task (VPC). Eye movements (duration and number of fixations) toward the top and bottom halves of the composite faces were recorded in 3-month-olds (Experiment 1) and adults (Experiment 3) using an eye-tracker system. Newborns were tested through a classical habituation paradigm in which the duration of visual fixations is coded by human expert observers. During the habituation-familiarization phase, all participants were presented with a top half of a face. The test phase involved the presentation of a pair of composite face stimuli. In one face stimulus of the pair, the top half of the face was identical to the one shown during familiarization, and the top of the other face was new.

Half of the participants were tested in the aligned condition, in which the left and right margins of the top and bottom face parts were properly aligned, and the other half were tested in the misaligned condition, in which the top and bottom face parts were offset horizontally.

Experiment 1

Method

Participants. Participants were twenty 3-month-old healthy and full-term infants (13 males, mean age = 99 days). They were middle-class infants and all of them except one were Caucasian. Seven additional infants were tested but not included in the final sample because they showed a strong position preference (i.e., they looked to one direction more than 90% of the time). Infants were tested only if awake and in an alert state after parents gave their informed consent.

Stimuli. The original stimuli were digitized, high-quality gray-scale images of 12 Caucasian female full-front faces, posing with a neutral expression and placed on a black background with no hair. All faces were unfamiliar to participants. Face images were split into two parts at the level of the middle of the nose, generating a top and a bottom segment. Composite faces were created by joining the top segment of one face to the bottom segment of a different face image (Young et al., 1987; Figure 1). By that matching, six pairs of faces were created. Within each pair, the bottom face half was the same. In the *aligned condition*, the top and the bottom face segments were properly horizontally aligned, whereas in the *misaligned condition* the top half of each face was shifted horizontally to the left, so that the middle of the nose in the bottom segment was positioned next to the extreme right side of the top segment. In both conditions, the two halves of each composite face were separated by a gap of 3 mm. When projected onto the screen, the top part presented in the habituation phase was 8 × 12 cm (about 8° × 11° of visual angle). Aligned faces were 15.5 × 12 cm (about 15° × 11° of visual angle). Misaligned faces were 15.5 × 17.5 cm (about 15° × 17° of visual angle). The distance between the center of the screen and the vertical median line of each stimulus was about 13.5 cm (about 13° of visual angle).

Apparatus. Infants were placed in an infant seat at a distance of about 60 cm from the computer screen where the stimuli were presented. A system for the automatic registration of eye movements made by Applied Science Laboratories (ASL) was employed and consisted of an infrared camera located at the bottom of the computer screen. Using a remote control and a live image of the participant's left eye, an experimenter, blind to the purpose of the study, guided the camera to keep the participant's eye constantly in focus. The eye-tracking system automatically detected the position of the pupil and the corneal reflection of the infrared light-emitting diodes (LEDs) in the eye. Because these signals changed as a function of the observer's gaze direction, the apparatus determined, with a frequency of 50 Hz, the x–y coordinates corresponding to the participant's fixation points during stimulus presentation. Two crosses of different colors corresponding to the signals coming from the participant's pupil and corneal reflection were superimposed on the images of the stimuli presented, giving a second experimenter a direct indication of the quality of the signal collected during the experiment. Stimulus presentation and data collection were performed using E-Prime 1.1. (Psychological Software Tool, Pittsburgh, PA).

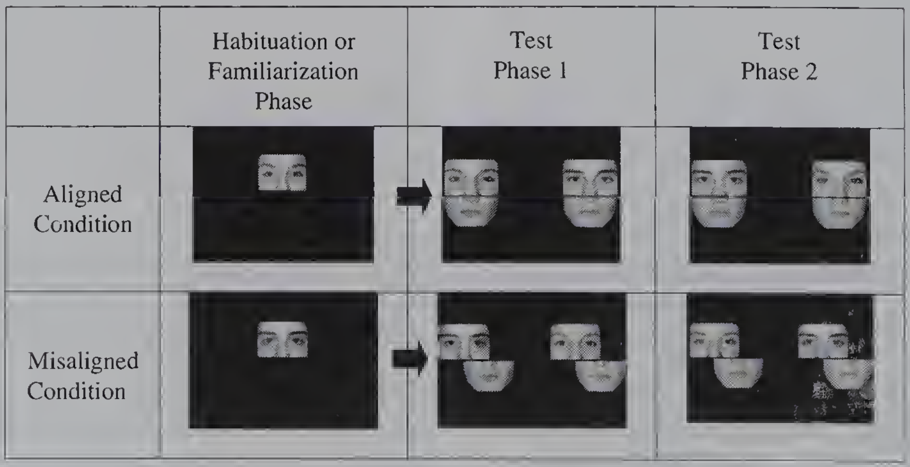


Figure 1. The stimuli and procedure used in Experiments 1–3.

Procedure. Each infant was tested on a single occasion in a quiet and dark room, seated approximately 60 cm from a 19-in. (1,024 × 768 pixels) computer screen. The experiment started with a calibration phase immediately followed by the habituation phase and the test phase. During calibration, a smiley face cartoon was presented in the center of the screen. When the infant started to look at it, the smiley moved to the top left corner of the screen and remained in this position until the infant fixated it. Then, it moved to the bottom right corner and remained in this position. These three positions were used to compute the pupil-corneal reflection from three points on the screen, allowing the system to derive gaze direction during the habituation and the test phases. Calibration accuracy was checked and repeated if necessary.

An infant-control habituation paradigm (Slater, Morison, & Rose, 1985) was used. Applying the ASL algorithm, a fixation was defined as a period of a least 100 ms during which the fixation point did not change by more than 1° of visual angle. During the habituation phase, the top half portion of a face was presented at the center of the screen. The infant was judged to have habituated when, from the fourth fixation on, the sum of any three consecutive fixations was 50% or less than the total of the first three. In the test phase, each infant was given two paired presentations of the test stimuli, in which the left–right stimulus position was reversed. In one face of the pair the top half was identical to that shown during habituation, in the other face the top half was new. Both top halves were matched with the same bottom half of a third new face. Each presentation lasted when a total of 10 s of looking at the novel and/or familiar stimuli had been accumulated. Half of the infants were tested in the aligned condition, the other half were tested in the misaligned condition.

Number of fixations and total fixation times were calculated by computing the sum of all fixations within four areas of interest (AOI) of the composite faces presented in the test phase (AOIs, Figure 2), which corresponded to the face halves that appeared on the left and right sides of the screen: AOI 1 and AOI 2 for the top halves of faces presented, respectively, on the left and on the right of the screen; AOI 3 and AOI 4 for the bottom halves of faces presented on the left and on the right of the screen.

Results

All infants reached the habituation criterion. A paired-sample t test was run to compare total fixation times to reach the habituation criterion in each condition, aligned condition ($M = 27{,}180$ ms, $SE = 5{,}755$) versus misaligned condition ($M = 19{,}618$ ms, $SE = 3{,}553$), $t(18) = 1.12$, *ns*. The difference was not significant.

To understand if infants' gaze behavior differed depending on the alignment versus misalignment of the top and bottom face portions, total looking times and number of fixations toward the top and bottom parts of the faces in the test phase were analyzed. The face that comprised the top half presented in the habituation phase was considered as familiar whereas the other was considered as novel. Preliminary analyses of variance (ANOVAs) on number of fixations and total fixation times toward the stimuli revealed that no main effect or interactions involved the factor presentation (first vs. second). Therefore, data were collapsed across this factor in two distinct three-way ANOVAs, performed on total fixation times and number of fixations, with test (novel vs. familiar) and face half (top vs. bottom) as within-subjects factors and testing condition (aligned vs. misaligned) as between-subjects factor.

Results of the ANOVA on total fixation times showed a main effect of the face half factor, $F(1, 18) = 82.53$, $p < .001$, $\eta^2 = .82$. Infants looked longer at the top face part ($M = 3{,}382$ ms, $SE = 206$) than at the bottom face part ($M = 720$ ms, $SE = 159$).

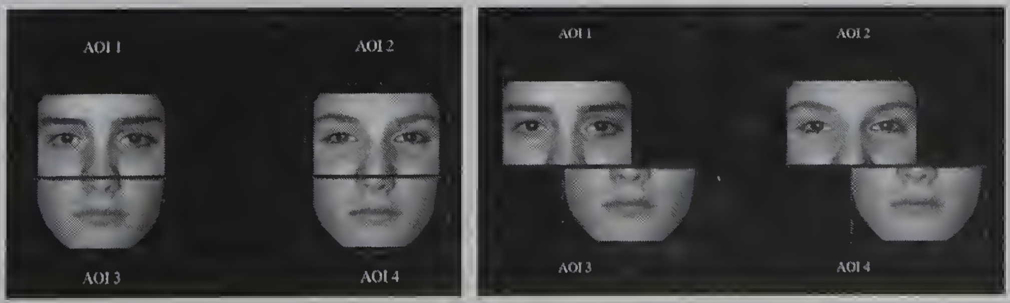


Figure 2. Areas of interest (AOIs) used in Experiments 2 and 3.

Furthermore, the main effect of test indicated that the familiar face (M = 2,322 ms, SE = 196) was fixated longer than the novel face (M = 1,780 ms, SE = 110), $F(1, 18) = 5.72$, $p < .05$, $\eta^2 = .24$. This test main effect could be explained in relation to the Test × Testing Condition significant interaction, $F(1, 18) = 4.56$, $p < .05$, $\eta^2 = .20$, which qualifies the main effect (Figure 3). Indeed, planned t tests showed that, in the misaligned test condition, infants looked longer at the familiar (M = 5,326 ms, SE = 645) than at the novel (M = 3,273 ms, SE = 366) composite stimulus, $t(9) = 2.66$, $p < .05$. On the contrary, in the aligned test condition, looking time toward the novel (M = 3,847 ms, SE = 244) and the familiar (M = 3,963 ms, SE = 446) faces did not differ, $t(9) = 0.24$, *ns*.

Results of the ANOVA on number of fixations showed a main effect of face half, $F(1, 18) = 88.59$, $p < .001$, $\eta^2 = .83$, and a significant Face Half × Testing Condition interaction, $F(1, 18) = 4.55$, $p < .05$, $\eta^2 = .20$, showing that the number of fixations infants oriented to the top face part (M = 40, SE = 2) was greater than the number of fixations toward the bottom face part (M = 11 ms, SE = 1), particularly in the misaligned (M = 44, SE = 3 vs. M = 9, SE = 3) as compared to the aligned test condition (M = 35, SE = 1 vs. M = 13, SE = 2).

Discussion

In accordance with previous evidence (Gallay, Baudouin, Durand, Lemoine, & Lécuyer, 2006; Turati, Valenza, Leo, & Simion, 2005), our findings showed that in both the aligned and the misaligned conditions, infants' visual exploration was focused on the top portion of the face, that comprised the eye region. More interestingly, results revealed a discrepancy in infants' looking times according to the alignment or misalignment of the faces. Three-month-olds looked longer at the face with a familiar top half when the top and the bottom face halves were misaligned, thus revealing that they discriminated the two stimuli. Conversely, in the aligned condition infants looked equally long to the familiar and novel stimuli, indicating that they did not differentiate between the two. This suggests that the bottom part of the composite faces affected infants' performance in both the aligned and the misaligned conditions. In the aligned condition, the novel bottom face halves impaired infants' discrimination of the two composite faces. In the misaligned condition, the addition of the bottom face half determined a persistent and extensive reexploration of the familiar top face half. Indeed, previous studies that tested infants' ability to recognize a learned face over strong modifications—such as partial occlusion, or rotation—obtained a familiarity preference (Gava, Valenza, Turati, & de Schonen, 2008; Walton, Amstrong, & Bower, 1997; see also Sirios & Mareschal, 2002, 2004). In the current experiment, infants in the misaligned condition not only recognized the familiarity of the top portion of the configuration but also detected that a novel never-seen bottom face part was added to the familiar part. This induced infants' extensive reexploration of the whole partially familiar stimulus.

The alignment versus misalignment of faces also affected the trajectories of infants' eye movements, as infants' tendency to fixate the top versus the bottom face halves was more evident in the misaligned condition, that is when the top part of the faces are disjoined from the bottom parts. Thus, overall, evidence obtained through the use of an eye-movement recording VPC procedure supports the existence of the composite effect in 3-month-old infants.

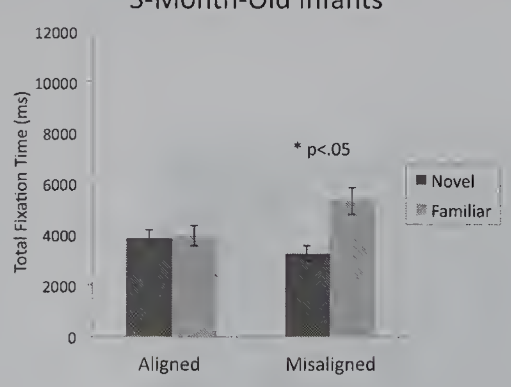


Figure 3. Total fixation time at the novel and the familiar stimuli as a function of test condition (aligned vs. misaligned condition) in Experiment 1.

Experiment 2

Method

Participants. Participants were 28 (14 females) healthy, full-term newborns (mean age = 45 hr), recruited at the maternity ward of the Pediatric Clinic of the University of Padova. All of them were middle-class infants and Caucasian. Nine additional newborns were tested but were not included in the final sample because 1 showed a strong position preference (i.e., he looked to one direction more than 80% of the time) and 8 did not complete testing due to fussiness. All of them met the screening criteria of normal delivery, a birth weight between 2,460 and 4,100 g, and an Apgar score between 8 and 10 at 5 min. Newborns were tested only if they were awake and in an alert state, after parents gave their informed consent.

Stimuli. The stimuli were identical to those presented in Experiment 1, except for dimensions expressed in visual angles that changed as 3-month-old infants were placed at a greater distance (60 cm) from the screen than newborns (30 cm). When projected on the screen, the two top parts presented in the habituation phase were 8 × 12 cm (about 15° × 23° of visual angle). Aligned faces were 15.5 × 12 cm (about 27° × 23°). Misaligned faces were 15.5 × 17.5 cm (about 27° × 33°). The distance between the center of the screen and the vertical median line of each figure was about 13.5 cm (about 26°).

Apparatus. The infant sat on the experimenter's lap, 30 cm from a 30-in. computer screen (2,560 × 1,600). Infants' eyes were aligned with a red flickering LED at the center of the screen, which was used to attract the infant's gaze at the start of each trial and to check that the infant's sight was level with the horizontal midline of the screen during the testing session. Stimuli were projected bilaterally on a black screen at a distance of approximately 7.5 cm (15°) from a central fixation point. A video camera, mounted above the central screen, recorded the infant's eyes movements. To prevent interference from irrelevant distracters, plain white curtains were drawn on both sides of the chair where the experimenter sat with the baby in her lap.

Procedure. Testing began as soon as the newborn looked at the central flickering LED. The experiment was carried out using an infant-control habituation procedure (Horowitz, Paden, Bhana, & Self, 1972). At present, the eye-tracker system is inappropriate for use with newborns because of the poor differentiation of the pupil at birth, which renders almost impossible for the infrared camera to detect the position of the pupil and the corneal reflection of the infrared LEDs in the newborns' eye. An experimenter, who was naïve to the hypotheses under investigation, watched the infants' eyes by means of a video monitor system and coded the duration of newborns' fixations toward the novel or the familiar composite face by pressing a key on the computer keyboard. During the habituation phase, newborns were presented with two top halves of the same face presented bilaterally. Bilateral rather than central presentation was chosen because it is difficult for an observer to decide if newborns are actually looking at a centrally presented stimulus or if they simply do not move their eyes from the central position. Also, at birth photoreceptors in the central fovea are very immature, thus resulting in a poor vision in the central area of visual field (Abramov et al., 1982; Atkinson & Braddick, 1989).

The stimuli remained on the screen until the habituation criterion was established through the recording of the duration of individual fixations. The duration of each fixation on the stimulus was recorded by pressing a button that was connected to the computer. During the habituation phase, the same stimulus was presented on the left and on the right, so the amount of looking was recorded irrespective of the side. The habituation criterion was identical to the one used in Experiment 1. When the habituation criterion was reached, the stimuli were automatically turned off and the central flickering LED was turned on. When the infants' gaze was realigned to the central LED, a preference test phase started. Each infant was given two paired presentations of the stimuli. During each presentation, newborns were simultaneously presented with a pair of composite faces. In one face of the pair, the top half was identical to that shown during habituation, in the other face the top half was new. Both top halves were matched with the same bottom half of a third new face. The two-paired stimuli were always shown in both left and right positions, the position being reversed from Presentation 1 to Presentation 2. During the preference test phase, the experimenter recorded the duration of infant's fixations and the number of fixations on each stimulus by pressing two different push buttons depending on whether the infant looked at the right or the left position. Each presentation lasted until each stimulus had been fixated at least once and a total of 20 s of looking had been accumulated. The test presentations were kept longer in Experiment 2 (20 s of fixation time) than in Experiment 1 (10 s of fixation time) to give newborns more time to process the stimuli and eventually manifest a novelty preference. Half of the infants were tested in the aligned condition, the other half were tested in the misaligned condition.

Results

All newborns reached the habituation criterion. A paired-sample t test was run to compare total fixation times to reach the habituation criterion in each condition (aligned condition, M = 65,210 ms, SE = 10,456 vs. misaligned condition, M = 74,886 ms, *SE* = 11,220). The comparison did not reach statistical significance, *t*(27) = 1.23, *ns*.

Preliminary ANOVAs on total fixation times and number of fixations toward the stimuli revealed that no main effect or interactions involved the factor presentation (first vs. second). Therefore, data were subsequently collapsed in two ANOVAs with testing condition (aligned vs. misaligned) as between-subjects factor and test (novel vs. familiar) as within-subjects factors, one for each dependent variable. The face stimulus that comprised the top half presented in the habituation phase was considered as familiar. There was no main effect, *F*(1, 26) = 0.45, *ns*, or interactions involving the factor test, *F*(1, 26) = 0.99, *ns* (Figure 4). Results showed that newborns looked equally long at the novel stimulus as compared to the familiar stimulus in the both the aligned (*M* = 22,248 ms, *SE* = 5,458 vs. *M* = 22,494 ms, *SE* = 8,131) and misaligned (*M* = 20,542 ms, *SE* = 5,834 vs. *M* = 20,733 ms, *SE* = 7,745) conditions. Furthermore, analyses with total number of fixations showed no main effect of the factor test, *F*(1, 26) = 0.13, *ns*, nor an interaction between the factors test and testing condition, *F*(1, 26) = 1.85, *ns*.

Discussion

The results of Experiment 2 showed that neither newborns tested in the aligned condition nor newborns tested in the misaligned condition discriminated the familiar from the novel composite face. Different possible explanations of the obtained results might be provided. One possibility is that newborns based their response in the test phase on the bottom half of the composite faces, which was identical in both the stimuli that comprised the novel and the familiar top halves. This explanation is not plausible given that less salient and contrasted areas are embedded in the bottom as compared with the top face part (Gava et al., 2008; Turati, 2004). Alternatively, newborns may have reacted to the significant perceptual change that occurred between the habituation phase (in which only one top half of a face was presented) and the test phase (in which two faces were presented, each comprising a novel bottom half) and showed a persistent and comparable exploration of both the familiar and the novel composite face. Finally, newborns might be unable to recognize a portion of a face—that is, only the top part rather than the full face—given that first-order face information is absent and second-order face information is poorly represented, the spatial relations between eyes and nose and eyes and mouth being disrupted. Future studies might further explore this issue by presenting newborns with a composite face rather than a top face part, in the habituation phase.

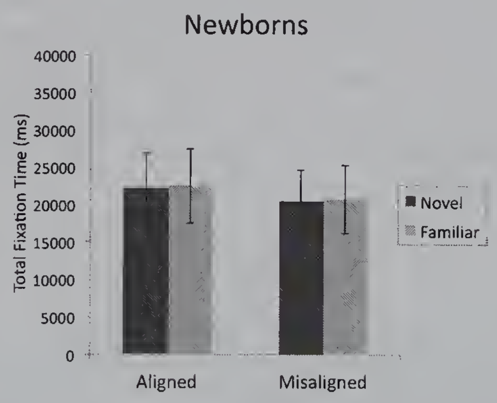


Figure 4. Total fixation time at the novel and the familiar stimuli as a function of test condition (aligned vs. misaligned condition) in Experiment 2.

Experiment 3

To ground the results obtained in Experiments 1 and 2 in an adult model, Experiment 3 tested a group of adults in a similar VPC task. In adults, the active control of eye fixations plays an important functional role in a wide variety of cognitive and perceptual tasks (for a review, see Henderson, 2006). Adults' eye fixations during face perception and recognition are predominantly directed to internal facial features, particularly the eyes, then nose and mouth (Althoff & Cohen, 1999; Henderson, Falk, Minut, Dyer, & Mahadevan, 2001; Stacey, Walker, & Underwood, 2005). Recently, a functional relation between eye movements and face perception has been demonstrated, in that a deficit in recognition emerged when participants were required to keep their eyes fixated in one central position (i.e., restricted viewing condition) rather than to move their eyes freely (i.e., free-viewing condition; Henderson, Williams, & Falk, 2005). Finally, Williams and Henderson (2007) found that the same features of a face were fixated regardless of whether the face was right side up or upside down and interpreted their findings as indicating that the face inversion effect is not the result of a different pattern of eye movements (but see Gallay et al., 2006, for qualitative differences in visual exploration strategies between upright and inverted faces in 4-month-olds).

Through the recording of adults' eye movement during a composite task, Experiment 3 was aimed at test gaze behavior during holistic face processing. Despite numerous studies on adults' composite face illusion, only a single study recorded eye movements while adults performed a delayed matching task of top halves of faces aligned or misaligned with bottom face halves (de Heering et al., 2008). Results showed that participants fixated longer the top parts of aligned than misaligned faces, suggesting the presence of a difference in eye-movement strategies between the aligned and misaligned conditions. However, manual reaction time data were strictly linked to eye-recording data, as suggested by the high correlations between participants' manual response time and ocular fixation time in aligned and misaligned conditions. When the effect of participants' manual response time was controlled, considering the variable as a covariate, the difference in mean fixation time on the top parts of aligned and misaligned faces disappeared, suggesting that longer fixation time on the top parts of aligned versus misaligned faces was due to the time devoted to respond manually at aligned versus misaligned faces, aligned faces being processed more slowly.

In light of this result, in the current experiment adults' eye movements were recorded in a free-viewing task that does not require a manual response. In parallel with the experimental paradigm used with 3-month-old and newborn infants in Experiments 1 and 2 and with studies that used the VPC task with adults (Richmond, Sowerby, Colombo, & Hayne, 2004), participants were familiarized with the top half of a face in a single fixed 15-s presentation. In the subsequent two fixed 10-s presentation test phases, a pair of composite faces was shown in which one of the face top halves was identical to the top half of the face presented during familiarization and the other top half was new. Half of the participants were tested in the aligned condition, and the other half were tested in the misaligned condition, so that alignment of the composite stimuli was varied between subject as in Experiments 1 and 2. No instructions were given, except that participants had to watch the images appearing on the computer screen moving their eyes freely.

Method

Participants. The participants were 22 undergraduate middle-class, Caucasian students (17 females, $M = 22.8$ years) from the Department of Psychology at the University of Padua (Italy). Seven additional participants were tested but excluded from the final sample due to failure to reach criteria established for data analyses (their fixation time toward stimuli was less than 70% of the total presentation time during both first and second test phases). All of them had normal or corrected to normal vision.

Stimuli. The stimuli were identical to those used in Experiments 1 and 2.

Apparatus and procedure. The apparatus and the procedure were the same as with 3-month-old infants in the previous experiment, with one exception. Following the calibration phase, instead of the habituation phase, participants were familiarized with a top half of a face projected in the center of the screen for a fixed presentation period of 15 s. Half of the participants were tested in the aligned condition, whereas the other half in the misaligned condition. Participants were given no instructions except that they had to watch the images appearing on the screen.

Results

To compare total fixation times in the aligned and misaligned conditions during the familiarization phase, a paired-sample t test was run (aligned condition, $M = 11{,}773$ ms, $SE = 762$ vs. misaligned condition, $M = 12{,}709$ ms, $SE = 434$), $t(20) = 1.06$, *ns*. As expected, the difference was not significant.

As with infants in Experiment 1, the aims of the statistical analyses performed on the data were to understand (a) whether adults' ability to discriminate the familiar from the novel composite face varied as a function of the alignment or misalignment of the face halves, and (b) whether the pattern of adults' fixation durations to the top and bottom halves of the composite faces varied between the two alignment conditions.

The face stimulus that comprised the top half presented in the habituation phase was considered familiar whereas the other was considered novel. Two distinct mixed-model ANOVAs were performed on total fixation times and number of fixations with presentation (first vs. second), test (novel vs. familiar), and face half (top vs. bottom) as within-subjects factors and testing condition (aligned vs. misaligned) as between-subject factor.

Results of the ANOVA on total fixation times showed that adults looked longer at the top face halves ($M = 2{,}869$ ms, $SE = 121$) than at the bottom face halves ($M = 1{,}062$ ms, $SE = 107$) as showed by the main effect of face half, $F(1, 20) = 65.33$, $p < .001$, $\eta^2 = .76$. More importantly, there was a significant interaction between the factors test and testing condition, $F(1, 20) = 6.51$, $p < .05$, $\eta^2 = .24$ (Figure 5). Planned *t* tests showed that in the misaligned test condition, adults looked longer at the novel stimulus ($M = 9{,}107$ ms, $SE = 516$) than at the familiar stimulus ($M = 6{,}762$ ms, $SE = 498$), $t(10) = 2.41$, $p < .05$. On the contrary, in the aligned test condition, mean fixation time toward the novel face ($M = 7{,}229$ ms, $SE = 435$) and the familiar face ($M = 8{,}356$ ms, $SE = 539$) did not differ, $t(10) = 1.19$, *ns*. This suggests that, in the aligned condition, the bottom part of the composite faces affected adults' performance, impairing their face recognition ability.

Finally, the interaction between the factors presentation and test also reached statistical significance, $F(1, 20) = 9.55$, $p < .05$, $\eta^2 = .32$. Specifically, planned *t* tests showed that adults looked longer at the novel stimulus ($M = 4{,}556$ ms, $SE = 257$) than at the familiar one ($M = 3{,}320$ ms, $SE = 243$) during the first test phase, $t(21) = 2.52$, $p < .05$, whereas there were not differences in looking behavior toward novel ($M = 3{,}612$ ms, $SE = 231$) and familiar ($M = 4{,}239$ ms, $SE = 285$) stimulus in the second test phase, $t(21) = 1.24$, *ns*.

The ANOVA on number of fixations revealed no main effect or interactions involving the factor Presentation (first vs. second). Therefore, data were collapsed across this factor in a subsequent three-way ANOVA with test (novel vs. familiar) and face half (top vs. bottom) as within-subjects factors and testing condition (aligned vs. misaligned) as between-subjects factor.

Results indicated a main effect of face half, $F(1, 20) = 90.34$, $p < .001$, $\eta^2 = .82$. The number of fixations adults made to the top part of the face ($M = 36$, $SE = 2$) was greater than the number of fixations toward the bottom part of the face ($M = 13$, $SE = 1$). Furthermore, the significant interaction between face half and testing condition, $F(1, 20) = 5.38$, $p < .05$, $\eta^2 = .21$, showed that the difference between fixations to the top part of the face and the bottom part of the face was greater in the aligned test condition ($M = 39$, $SE = 2$ vs. $M = 10$, $SE = 1$) than in the misaligned test condition ($M = 34$, $SE = 2$ vs. $M = 17$, $SE = 3$). Furthermore, there was a significant interaction between the factors test and testing condition, $F(1, 20) = 6.23$, $p < .05$, $\eta^2 = .24$. Planned *t* tests showed that, in the misaligned test condition, the number of fixations toward the novel stimulus ($M = 29$, $SE = 2$) was greater than the number of fixations toward the familiar stimulus ($M = 22$, $SE = 2$), $t(10) = 2.97$, $p < .05$. On the contrary, in the aligned test condition, the number of fixations toward the novel ($M = 24$, $SE = 2$) and the familiar faces ($M = 25$, $SE = 1$) did not differ, $t(10) = 0.61$, *ns*.

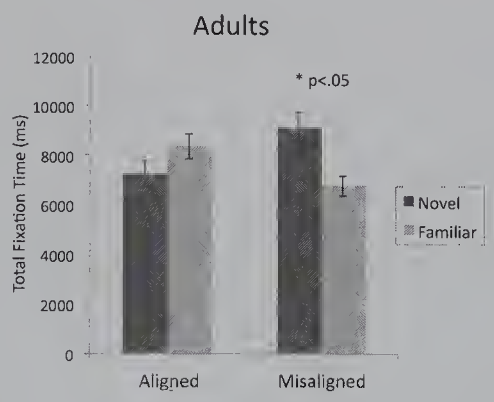


Figure 5. Total fixation time at the novel and the familiar stimuli as a function of test condition (aligned vs. misaligned condition) in Experiment 3.

Discussion

Evidence from the present experiment appears in line with previous findings showing that adults' fixations toward faces are directed predominantly toward the eye regions (e.g., Althoff & Cohen, 1999; Henderson et al., 2001; Henderson et al., 2005). Importantly, adults looked longer and directed a greater number of fixations toward the face with a novel top half in the misaligned condition. Conversely, in the aligned condition adults looked equally long and oriented their gaze approximately the same number of times toward the novel and familiar stimuli. This suggests that, in the aligned condition, the bottom part of the composite faces affected adults' performance, demonstrating the presence of the composite face effect. Aligned versus misaligned stimulus conditions elicited differential looking times and numbers of fixations for the novel and the familiar composite face, providing evidence that eye movements may be affected by the encoding of holistic face information.

Overall, the present work is consistent with the idea that eye movements during face recognition are functional (Henderson et al., 2005) but is partially in contrast with the findings reported by de Heering et al. (2008) on the composite effect and by Williams and Henderson (2007) on the inversion effect, according to which eye-movement patterns are not able to reflect holistic processing strategies. The discrepancies between our studies and previous work may be explained by at least three major methodological differences. First, in our study participants were not required to provide a manual response. This may have avoided high correlations between participants' manual response time and ocular fixation time that could produce distortions in the results related to eye movements, as longer manual reaction times are associated with longer fixations. Second, with respect to the study by de Heering et al., participants were not explicitly required to look at the top part of a face in order to recognize it but were allowed to freely move their eyes. In accord with de Heering et al., we found that the top halves of the composite faces in the test phase were fixated longer than the bottom halves, irrespective of whether the face halves were aligned or misaligned. However, different from de Heering et al., in our study the aligned and misaligned conditions affected the amount of visual exploration of the novel and the familiar composite faces and number of fixations directed toward the top and the bottom composite face part. Third, differently from the study by Williams and Henderson, the use of the composite paradigm allowed us to split faces into only two critical AOI (top face part and bottom face part) rather than in multiple areas related to each face feature (left eye, right eye, nose, mouth, chin, forehead, etc.). This might have produced a more adequate measure of holistic, rather than feature-based, face processing strategies.

General Discussion

The main outcome of this study is that for the first time a clear demonstration was provided that young infants are capable of processing a face holistically, as an undifferentiated unit in which face parts and their relations are glued together (Experiment 2). The present work extends previous findings that reported a measure of the composite effect at 3 years of age (Macchi Cassia et al., 2009) and of the part–whole effect at 4 years of age (Pellicano & Rhodes, 2003). Also, the present results extend previous findings obtained by Cohen and colleagues, demonstrating that infants are capable of detecting correlations not only between outer and inner face features (Cashon & Cohen, 2001, 2003) but also between the sole inner features of a face.

Importantly, this study utilized a version of the composite paradigm adapted to infants but comparable to the measures employed to examine holistic processing in older children and adults. In fact, this paradigm was shown to be a potential measure of the composite face effect also in adults (Experiment 3), demonstrating that eye movements (duration and number of fixations) may be sensitive to holistic face processing during the entire life span. The use of the same task throughout development, from early infancy to adulthood, allows researchers to highlight both differences and similarities in the visual behavior across age groups. First, results clearly show that the facial features that attract both adults' and infants' ocular fixations are almost the same, in that at both ages a clear tendency to explore the top rather than the bottom face part was found. Second, although both age groups visually recognized the familiar stimulus in the misaligned but not in the aligned condition, infants preferred looking at the face with the familiar top part whereas adults tended to look longer at the face with the novel top part. Thus, adults did not need to reexplore the stimulus that comprises the familiar top of the face and directly focused attention on the face stimulus with a novel top. Overall, evidence from this study reveals that the tuning toward configural information appears very early in life, although gradual experience-based developmental processes will progressively refine early configural abilities.

References

Abramov, I., Gordon, J., Hendrickson, A., Hainline, L., Dobson, V., & LaBossiere, E. (1982). The retina of the newborn human infant. *Science*, *217*, 265–267.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 997–1010.

Atkinson, J., & Braddick, O. (1989). Development of basic visual functions. In A. Slater & G. Bremner (Eds.), *Infant development* (pp. 7–41). Hove, UK: Erlbaum.

Bartrip, J., Morton, J., & de Schonen, S. (2001). Responses to mother's face in 3-week to 5-month-old infants. *British Journal of Developmental Psychology*, *19*, 219–232.

Bertin, E., & Bhatt, R. S. (2004). The Thatcher illusion and face processing in infancy. *Developmental Science*, *7*, 431–436.

Bhatt, R. S., Bertin, E., Hayden, A., & Reed, A. (2005). Face processing in infancy: Developmental changes in the use of different kinds of relational information. *Child Development*, *76*, 169–181.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, *1*, 253–274.

Cashon, C. H., & Cohen, L. B. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant and Child Development*, *10*, 83–92.

Cashon, C. H., & Cohen, L. (2003). The construction, deconstruction, and reconstruction of infant face perception. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood:*

Current perspectives (pp. 13–25). New York: Nova Science.

de Haan, M., Johnson, M. H., Maurer, D., & Perrett, D. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognitive Development*, *1*, 659–678.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology*, *96*, 57–70.

de Heering, A., Rossion, B., Turati, C., & Simion, F. (2008). Holistic face processing can be independent of gaze behaviour: Evidence from the composite face illusion. *Journal of Neuropsychology*, *1*, 183–195.

Deruelle, C., & de Schonen, S. (1998). Do the right and left hemispheres attend to the same visuospatial information within a face in infancy? *Developmental Neuropsychology*, *14*, 535–554.

Gallay, M., Baudouin, J.-Y., Durand, K., Lemoine, C., & Lécuyer, R. (2006). Qualitative differences in the exploration of upright and upside-down faces in four-mouth-old infants: An eye-movement study. *Child Development*, *77*, 984–996.

Gava, L., Valenza, E., Turati, C., & de Schonen, S. (2008). Effect of partial occlusion on newborns' face preference and recognition. *Developmental Science*, *11*, 563–574.

Hayden, A., Bhatt, R. S., Reed, A., Corbly, C. R., & Jospeh, J. E. (2007). The development of expert face processing: Are infants sensitive to normal differences in second-order relational information? *Journal of Experimental Child Psychology*, *3*, 85–98.

Henderson, J. M. (2006). Eye movements. In C. Senior, T. Russell, & M. Gazzaniga (Eds.), *Methods in mind* (pp. 171–191). Cambridge, MA: MIT Press.

Henderson, J. M., Falk, R. J., Minut, S., Dyer, F. C., & Mahadevan, S. (2001). Gaze control for face learning and recognition by humans and machines. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Segmentation processes in vision* (pp. 463–482). Amsterdam: Elsevier.

Henderson, J. M., Williams, C. C., & Falk, R. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*, 98–106.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, *23*, 65–74.

Horowitz, F. D., Paden, L., Bhana, K., & Self, P. (1972). An infant-control procedure for studying infant visual fixation. *Developmental Psychology*, *7*, 90.

Leo, I., & Simion, F. (2009). Face processing at birth: A Thatcher illusion study. *Developmental Science*, *12*, 492–498.

Macchi Cassia, V., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science*, *12*, 236–248.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255–260.

Mondloch, C. J., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite effect in six-year-old children: Evidence for adult-like holistic face processing. *Visual Cognition*, *15*, 564–577.

Morton, J., & Johnson, M. H.. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, *98*, 164–181.

Pascalis, O., de Haan, M., Nelson, C. A., & de Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 249–260.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science*, *14*, 618–622.

Richmond, J., Sowerby, P., Colombo, M., & Hayne, H. (2004). The effect of familiarization time, retention interval, and context change on adults' performance in the visual paired-comparison task. *Developmental Psychobiology*, *44*, 146–155.

Sirios, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*, 293–298.

Sirios, S., & Mareschal, D. (2004). An interacting systems model of infant habituation. *Journal of Cognitive Neuroscience*, *16*, 1352–1362.

Slater, A., Bremner, G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. (2000). Newborn infants' preference for attractive faces: The role of internal and external facial features. *Infancy*, *1*, 265–274.

Slater, A., Morison, V., & Rose, D. (1985). Pattern preferences at birth and their interaction with habituation-induced novelty preferences. *Journal of Experimental Child Psychology*, *39*, 37–54.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology*, *96*, 407–422.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *46*, 225–245.

Thompson, L. A., Madrid, V., Westbrook, S., & Johnston, V. (2001). Infants attend to second-order relational properties of faces. *Psychonomic Bulletin and Review*, *8*, 769–777.

Turati, C. (2004). Why faces are not special to newborns: An alternative account of the face preference. *Current Directions in Psychological Science*, *13*, 5–8.

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development*, *77*, 297–311.

Turati, C., Sangrigoli, S., Ruel, J., & de Schonen, S. (2004). Evidence of the face inversion effect in 4-month-old infants. *Infancy*, *6*, 275–297.

Turati, C., Valenza, E., Leo, I., & Simion, F. (2005). Three-month-olds' visual preference for faces and its underlying mechanisms. *Journal of Experimental Child Psychology*, *90*, 255–273.

Valenza, E., Simion, F., Macchi Cassia, V., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 892–903.

Walton, G. E., Amstrong, E. S., & Bower, T. G. R. (1997). Faces as forms in the world of the newborns. *Infant Behavior and Development*, *20*, 537–543.

Williams, C., & Henderson, J. M. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition*, *35*, 1977–1985.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1906–1910

Manuscripts Accepted for Publication

An Ecological Approach to Promoting Early Adolescent Mental Health and Social Adaptation: Family Centered Intervention in Public Middle Schools, Elizabeth Stormshak, Arin M. Connell, Marie-Helene Veronneau, Michael W. Myers, Thomas J. Dishion, Kathryn Kavanagh, and Allison S. Caruthers.

Assessing the Effects of a Work-Based Antipoverty Program for Parents on Youth's Future Orientation and Employment Experiences, Rachel Kaplan, Vonnie C. McLoyd, Kelly Purtell, and Aletha Huston.

Adolescents' Development of Skills for Agency in Youth Programs: Learning to Think Strategically, Reed Larson and Rachel Angus.

Intergenerational Continuity in Child Maltreatment: Mediating Mechanisms and Implications for Prevention, Lisa Berlin, Karen Appleyard, and Kenneth Dodge.

Two-Year-Olds Use the Generic/Non-Generic Distinction to Guide Their Inferences About Novel Kinds, Susan Graham, Samantha Nayer, and Susan Gelman.

Understanding Bullying and Victimization During Childhood and Adolescence: A Mixed Methods Study, Nancy Guerra, Kirk R. Williams, and Shelly Sadek.

A Large-Scale Evaluation of the KiVa Anti-Bullying Program, Antti Kärnä, Marinus Voeten, Elisa Poskiparta, Todd D. Little, Anne Kaljonen, and Christina Salmivalli.

Danish Majority Children's Reasoning About Exclusion Based on Gender and Ethnicity, Signe Juhl Moller and Harriet Tenenbaum.

How Does It Look? Level 2 Perspective-Taking at 36 Months of Age, Henrike Moll and Andrew Meltzoff.

Emotion Discourse, Social Cognition, and Social Skills in Children With and Without Developmental Delays, Rachel M. Fenning, Bruce L. Baker, and Jaana Juvonen.

A Monozygotic Twin Differences Study of Friends' Aggression and Children's Adjustment Problems, Frank Vitaro, Mara Brendgen, Michel Boivin, Stephane Cantin, Ginette Dionne, Richard E. Tremblay, Alain Girard, and Daniel Perusse.

Gender Differences in Adolescents' Daily Interpersonal Events and Well-Being, Lisa Flook.

Teacher–Child Relationships and the Development of Academic and Behavioral Skills During Elementary School: A Within- and Between-Child Analysis, Carolina Maldonado-Carreño and Elizabeth Votruba-Drzal.

The Effects of the Fast Track Preventive Intervention on the Development of Conduct Disorder Across Childhood, Conduct Problems Prevention Research group.

Object Similarity Bootstraps Young Children to Action-Based Verb Extensions, Etsuko Haryu, Mutsumi Imai, and Hiroyuki Okada.

The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions, Joseph Durlak, Roger Weissberg, Allison Dymnicki, Rebecca Taylor, and Kriston Schellinger.

Science Does Not Speak for Itself: Translating Child Development Research for the Public and its Policymakers, Jack Shonkoff and Susan Nall Bales.

Domain General Constraints on Statistical Learning, Erik Thiessen.

Reciprocal Relations Between Parenting and Adjustment in a Sample of Juvenile Offenders, Lela Rankin Williams and Laurence Steinberg.

A Randomized Controlled Trial of Child FIRST: A Comprehensive Home-Based Intervention Translating Research Into Early Childhood Practice, Darcy I. Lowell, Alice S. Carter, Leandra Godoy, Belinda Paulicin, and Margaret J. Briggs-Gowan.

Latino Adolescents' Experiences of Discrimination Across the First Two Years of High School: Correlates and Influences on Educational Outcomes, Aprile Benner and Sandra Graham.

The Generic/Non-Generic Distinction Influences How Children Interpret New Information About Social Others, Andrei Cimpian and Ellen M. Markman.

Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research, Stephanie Jones, Joshua Brown, and J. Lawrence Aber.

Conditional Reasoning With False Premises Facilitates the Transition Between Familiar and Abstract Reasoning, Henry Markovits and Hugues Lortie Forgues.

When Easy Comes Hard: The Development of Adaptive Strategy Selection, Rui Mata, Bettina von Helversen, and Joerg Rieskmap.

Coherent Motion Sensitivity and Reading Development in the Transition From Pre-Reading to Reading Stage, Bart Boets, Maaike Vandermosten, Piers Cornelissen, Jan Wouters, and Pol Ghesquière.

Same-Ethnic Friendship Preferences Among German and Turkish Preadolescents Attending Ethnically Heterogeneous Schools, Philipp Jugert, Peter Noack, and Adam Rutland.

Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect, Chiara Turati, Elisa Di Giorgio, Lara Bardi, and Francesca Simion.

Math-Gender Stereotypes in Elementary-School Children, Dario Cvencek, Andrew N. Meltzoff, and Anthony G. Greenwald.

Reporter Discrepancies Among Parents, Adolescents, and Peers: Adolescent Attachment and Informant Depressive Symptoms as Explanatory Factors, Katherine B. Ehrlich, Jude Cassidy, and Matthew J. Dykas.

Urban and Rural Chinese Adolescents' Judgments and Reasoning About Personal and Group Jurisdiction, Charles Helwig, Shaogang Yang, Dingliang Tan, Chunqiong Liu, and Tiffany Shao.

Childhood Trauma and Psychiatric Disorders as Correlates of School Dropout in a National Sample of Young Adults, Michelle V. Porche, Lisa R. Fortuna, Julia Lin, and Margarita Alegria.

The Role of Peer Relationships in the Development of Early School-Age Externalizing Problems, Cristina Sturaro, Pol A. C. van Lier, Pim Cuijpers, and Hans M. Koot.

Relationship Between Language Experiences in Preschool Classrooms and Children's Kindergarten and Fourth Grade Language and Reading Abilities, David K. Dickinson and Michelle V. Porche.

Intergenerational Transmission of Adaptive Functioning: A Test of the Interactionist Model of SES and Human Development, Thomas Schofield, Monica Martin, Katherine Conger, Brent Donnellan, Tricia Neppl, and Rand Conger.

Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring, Carla Herrera, Jean Baldwin Grossman, Tina J. Kauh, and Jennifer McMaken.

Remission of Depression in Parents: Links to Healthy Functioning in Their Children, Judy Garber, Jeffrey A. Ciesla, Elizabeth McCauley, Guy Diamond, and Kelly A. Schloredt.

Double Jeopardy: Poorer Social-Emotional Outcomes for Children in the NICHD SECCYD Experiencing Home and Child Care Environments That Confer Risk, Sarah Watamura, Deborah Phillips, Taryn Morrissey, Kathleen McCartney, and Kristen Bub.

Imagining a Way Out of the Gravity Bias: Preschoolers Can Visualize the Solution to a Spatial Problem, Amy Joh, Vikram Jaswal, and Rachel Keen.

Sequential Progressions in a Theory of Mind Scale: Longitudinal Perspectives, Henry Wellman, Fuxi Fang, and Candida Peterson.

Can Maltreated Children Inhibit True and False Memories for Emotional Information?, Mark Howe, Sheree Toth, and Dante Cicchetti.

Parents' Involvement in Children's Learning in the United States and China: Implications for Children's Academic and Emotional Adjustment, Cecilia Cheung and Eva Pomerantz.

Four Children and One Toy: Chinese and Canadian Children Faced With Potential Conflict Over a Limited Resource, Doran French, Xinin Chen, Janet Chung, Miao Li, Dan Li, Huichang Chen, and Li Wang.

Influence of Eye Gaze on Spoken Word Processing: An ERP Study With Infants, Eugenio Parise, Andrea Handl, Letizia Palumbo, and Angela Friederici.

Infant and Maternal Sensitivity to Interpersonal Timing, Anne Henning and Tricia Striano.

Emotion Regulation Profiles, Temperament, and Adjustment Problems in Pre-adolescents, Maureen Zalewski, Liliana J. Lengua, Anna C. Wilson, Anika Trancik, and Alissa Bazinet.

Developmental Pathways to Integrated Social Skills: The Roles of Parenting and Early Intervention in Raising Healthy Children, Catherine Ayoub, Claire Vallotton, and Ann Mastergeorge.

Consequences of ''Minimal'' Group Affiliations in Children, Yarrow Dunham, Andrew Baron, and Susan Carey.

Revisiting the Impact of Part-Time Work on Adolescent Adjustment: Distinguishing Between Selection, Kathryn C. Monahan, Joanna Lee, and Laurence Steinberg.

Parents as Moderators of the Impact of School Norms and Peer Influences on Aggression in Middle School Students, Albert Farrell, David B. Henry, Sally Mays, and Michael E. Schoeny.

The ABC's of Family Mealtimes: Observational Lessons for Promoting Healthy Outcomes for Children With Persistent Asthma, Barbara Fiese, Marcia A. Winter, and Joanna Corra Botti.

Language Development in Internationally Adopted Children: A Special Case of Early Second Language Learning, Karine Gauthier and Fred Genesee.

Third-Party Social Interaction and Word Learning From Video, Kate O'Doherty, Georgene L. Troseth, Priya Shimpi, Elizabeth Goldenberg, Nameera Akhtar, and Megan M. Saylor.

Computer-Assisted Remedial Reading Intervention for School Beginners at Risk for Reading Disability, Nina Saine, Marja-Kristiina Lerkkanen, Timo Ahonen, Asko Tolvanen, and Heikki Lyytinen.

Desire for Increased Autonomy and Adolescents' Perceptions of Peer Autonomy: ''Everyone Else Can; Why Can't I?'', Christopher Daddis.

Almost Thinking Counterfactually: Children's Understanding of Close Counterfactuals, Sarah Beck and Carlie Guthrie.

CSRP's Impact on Low-Income Preschoolers' Pre-Academic Skills: Self-Regulation as a Mediating Mechanism, C. Cybele Raver, Stephanie M. Jones, Christine Li-Grining, Fuhua Zhai, Kristen Bub, and Emily Pressler.

Does God Make It Real? Children's Belief in Religious Stories From the Judeo-Christian Tradition, Victoria Cox Vaden and Jacqueline D. Woolley.

Changes in Early Adolescents' Sense of Responsibility to Their Parents in the United States and China: Implications for Their Academic Functioning, Eva Pomerantz, Lili Qin, Qian Wang, and Huichang Chen.

Two- to Eight-Month-Old Infants' Perception of Dynamic Auditory-Visual Spatial Co-Location, Gavin Bremner, Alan Slater, Scott Johnson, Uschi Mason, Jo Spring, and Maggie Bremner.

Longitudinal Associations Between Marital Instability and Child Sleep Problems Across Infancy and Toddlerhood in Adoptive Families, Anne Mannering, Gordon Harold, Leslie Leve, Katherine Shelton, Daniel Shaw, Rand Conger, Jenae Neiderhiser, Laura Scaramella, and David Reiss.

Promoting Effective Parenting Practices and Preventing Child Behavior Problems in School, Laurie Brotman, Esther Calzada, Keng-Yen Huang, Sharon Kingston, Spring Dawson-McClure, Dimitra Kamboukos, Amanda Rosenfelt, Ami Schwab, and Eva Petkova.

Age 26 Cost–Benefit Analysis of the Child–Parent Center Early Education Program, Arthur Reynolds, Judy Temple, Barry White, Suh-Ruu Ou, and Dylan Robertson.

Infantile Amnesia Across the Years: A 2-Year Follow-Up of Children's Earliest Memories, Carole Peterson, Kelly L. Warren, and Megan M. Short.

Hemoglobin, Growth, and Attention of Infants in Southern Ethiopia, Nicki Aubuchon-Endsley, Stephanie L. Grant, Getenesh Berhanu, David G. Thomas, Sarah Schrader, Devon Eldridge, Tay S. Kennedy, and Michael Hambidge.

Transition From Crawling to Walking and Infants' Actions With Objects and People, Lana Karasik, Catherine Tamis-LeMonda, and Karen Adolph.

The Genetic and Environmental Covariation Among Psychopathic Personality Traits and Reactive and Proactive Aggression in Childhood, Serena Bezdjian, Catherine Tuvblad, Adrian Raine, and Laura A. Baker.

Structure Mapping and Relational Language Support Children's Learning of Relational Categories, Dedre Gentner, Florencia K. Anggoro, and Raquel S. Klibanoff.

Narrative Elaboration and Participation: Two Dimensions of Maternal Elicitation Style, Gigliana Melzi, Adina R. Schick, and Joy L. Kennedy.

Does Felt Gender Compatibility Mediate Influences of Self-Perceived Gender Nonconformity on Early Adolescents' Psychosocial Well-Being?, Meenakshi Menon.

Theory of Mind and Social Interest in Zero Acquaintance Play Situations, Chris Moore, Sandra Bosacki, and Shannon Macgillivray.

Media as Social Partners: The Social Nature of Young Children's Learning From Screen Media, Rebekah Richert, Michael B. Robb, and Erin I. Smith.

Protecting Children From the Consequences of Divorce: A Longitudinal Study of the Effects of Parenting on Children's Coping Processes, Clorinda Velez, Sharlene Wolchik, Jenn-Yun Tein, and Irwin Sandler.

Preschoolers' Quarantining of Fantasy Stories, Rebekah Richert and Erin I. Smith.

Effect of Foster Care on Young Children's Language Learning, Jennifer Windsor, Joann Benigno, Christine Wing, Patrick Carroll, Sebastian Koga, Charles Nelson, Nathan Fox, and Charles Zeanah.

Why Don't Smart Teens Have Sex? A Behavioral Genetic Approach, Kathryn Paige Harden and Jane Mendle.

Preschoolers' Use of Morphosyntactic Cues to Identify Generic Sentences: Indefinite Singular Noun Phrases, Tense, and Aspect, Andrei Cimpian, Trent J. Meltzer, and Ellen M. Markman.

Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability, Michele Mazzocco, Lisa Feigenson, and Justin Halberda.

Imitation in Infancy: Rational or Motor Resonance?, Markus Paulus, Sabine Hunnius, Marlies Vissers, and Harold Bekkering.

Effects of Non-Maternal Care in the First Three Years on Children's Academic Skills and Behavioral Functioning in Childhood and Early Adolescence, Sara Jaffee, Carol Van Hulle, and Joseph Rodgers.

Shyness-Sensitivity and Unsociability in Rural Chinese Children: Relations with Social, School, and Psychological Adjustment, Xinyin Chen and Li Wang.

Age Differences in Strategic Planning as Indexed by the Tower of London, Dustin Albert and Laurence Steinberg.

Young Children's Learning and Transfer of Biological Information From Picture Books to Real Animals, Patricia Ganea, Lili Ma, and Judy DeLoache.

Characterizing Children's Expectations About Expertise and Incompetence: Halo or Pitchfork Effects?, Melissa Koenig and Vikram K. Jaswal.

Wherein Lies Children's Intergroup Bias? Egocentrism, Social Understanding, and Social Projection, Dominic Abrams.

Can Children With Autism Spectrum Disorders "Hear" a Speaking Face?, Julia Irwin, Lawrence Brancazio, Lauren Tornatore, and D.H. Whalen.

The Home Learning Environment Across the First Five Years: Associations With Children's Vocabulary and Emergent Literacy Skills at Pre-Kindergarten, Eileen T. Rodriguez and Catherine S. Tamis-LeMonda.

A Longitudinal Examination of African American Adolescents' Attributions About Achievement Outcomes, Akilah Swinton, Beth Kurtz-Costes, Stephanie Rowley, and Ndidi Okeke.

Ethnic Stigma, Academic Anxiety, and Intrinsic Motivation in Middle Childhood, Cari Gillen-O'Neel, Diane N. Ruble, and Andrew J. Fuligni.

Examining the Black-White Achievement Gap Among Low-Income Children: Using the NICHD Study of Early Child Care and Youth Development, Margaret Burchinal, Kathleen McCartney, Laurence Steinberg, Robert Crosnoe, Sarah L. Friedman, Vonnie McLoyd, and Robert Pianta.

A Longitudinal Study of Religious Identity and Participation During Adolescence, Anna Lopez, Virginia Huynh, and Andrew Fuligni.

Benefits of Practicing 4 = 2 + 2: Nontraditional Problem Formats Facilitate Children's Understanding of Mathematical Equivalence, Nicole M. McNeil, Emily R. Fyfe, Lori A. Petersen, April E. Dunwiddie, and Heather Brletic-Shipley.

Reciprocal Associations Between Family and Peer Conflict in Adolescents' Daily Lives, Grace Chung, Lisa Flook, and Andrew Fuligni.

The Development of Distrust, Kimberly Vanderbilt, David Liu, and Gail Heyman.

Developing Relationships, Being Cool, and Not Looking Like a Loser: Social Goal Orientation Predicts Children's Responses to Peer Aggression, Karen D. Rudolph, Jamie L. Abaied, Megan Flynn, Niwako Sugimura, and Anna Monica Agoston.

Nonlinear Growth Curves in Developmental Research, Kevin Grimm, Nilam Ram, and Fumiaki Hamagami.

Developmental Continuity in Theory of Mind: Speed and Accuracy of Belief-Desire Reasoning in Children and Adults, Ian Apperly, Frances Warren, Benjamin Andrews, Jay Grant, and Sophie Todd.

Children's Judgments of Emotion From Conflicting Cues in Speech: Why 6-Year-Olds Are so Inflexible, Matthew Waxer and J. Morton.

Social-Emotional Effects of Early Childhood Education Programs in Tulsa, William Gormley Jr., Deborah A. Phillips, Katie Newmark, Kate Perper, and Shirley Adelstein.

The Development of Uncertainty Monitoring in Early Childhood, Kristen E. Lyons and Simona Ghetti.

Critical Consciousness Development and Political Participation Among Marginalized Youth, Matthew Diemer and Cheng-Hsien Li.

The Moderating Capacity of Racial Identity Between Perceived Discrimination and Psychological Well-Being Over Time Among African American Youth, Eleanor Seaton, Enrique Neblett, Rachel D. Upton, Wizdom Powell Hammond, and Robert Sellers.

Antecedents and Outcomes of Joint Trajectories of Mother–Son Conflict and Warmth During Middle Childhood and Adolescence, Christopher Trentacosta, Michael Criss, Daniel Shaw, Eric Lacourse, Luke Hyde, and Thomas Dishion.

Parenting and Infant Difficulty: Testing a Mutual Exacerbation Hypothesis to Predict Early Onset Conduct Problems, Michael Lorber and Byron Egeland.

Vision for Action in Toddlers: The Posting Task, Sandra Y. Street, Karin H. James, Susan S. Jones, and Linda B. Smith.

The Gap Between Spanish-Speakers' Word Reading and Word Knowledge: A Longitudinal Study, Jeannette Mancilla-Martinez and Nonie K. Lesaux.

African American and European American Children in Diverse Elementary Classrooms: Social Integration, Social Status, and Social Behavior, Travis Wilson and Philip C. Rodkin.

Characterizing and Comparing the Friendships of Anxious-Solitary and Unsociable Preadolescents, Gary Ladd, Becky Kochenderfer-Ladd, Natalie D. Eggum, Karen Kochel, and Erin McConnell.

The Effect of Narrative Cues on Infants' Imitation from Television and Picture Books, Gabrielle Simcock, Kara Garrity, and Rachel Barr.

Social Groups and Children's Intergroup Attitudes: Can School Norms Moderate the Effects of Social Group Norms?, Drew Nesdale and Michael J. Lawson.

Differential Susceptibility to Adolescent Externalizing Trajectories: Examining the Interplay Between CHRM2 and Peer Group Antisocial Behavior, Shawn Latendresse, John Bates, Jackson Goodnight, Kenneth Dodge, Jennifer Lansford, Gregory Pettit, John Budde, Alison Goate, and Danielle Dick.

Newborn Irritability Moderates the Association Between Infant Attachment Security and Toddler Exploration and Sociability, Brandi Stupica, Laura J. Sherman, and Jude Cassidy.

Child Development, November/December 2010, Volume 81, Number 6, Pages 1911–1916

Editorial Acknowledgments

Over the past year, many developmental scientists, in addition to our Consulting Editors, have participated in the peer review process at *Child Development*. These individual efforts are, invariably by their very nature, anonymous. But at the end of the calendar year, we are delighted to be able to acknowledge these experts' important contributions. It is not an exaggeration to state that their input is as much embodied in the pages of *CD* as are those of our authors. I also want to take this opportunity to acknowledge the immeasurable contributions of Mary Dozier, Nancy G. Guerra, D. Geoffrey Hall, and Betsy Hoza, all of whom completed their terms as Associate Editor during this calendar year and have elected to step down from this post. The field has benefited enormously from their collective wisdom. And on a personal note, it has been a great privilege for me to work with these individuals. I am also honored and delighted to announce that Karen Bierman, Gil Diesendruck, Bonnie J. Leadbeater, and Glenn I. Roisman have accepted appointments as Associate Editors of the journal. I welcome their contributions to *Child Development* and to the field.

—*Jeffrey J. Lockman*

August 1, 2009–July 30, 2010

Maxim Abelev
Jennifer Ablow
Frances Aboud
Dominic Abrams
Brian Ackerman
Marilyn Adams
Scott Adler
Nameera Akhtar
Daisuke Akiba
Natacha Akshoomoff
Kristen Alexander
Joseph Allen
Melissa Allen
Patricia Aloise-Young
Ellen Altermatt
Dima Amso
Thomas Anders
Daniel Anderson
David Anderson
Glenda Andrews
Florencia Anggoro
Daniel Ansari
Mark Antoniou
Evan Apfelbaum
Ian Apperly
Doreen Arcus
Mary Arimond
David Arnold
Sudha Arunachalam
Kathryn Asbury
Steven Asher
Richard Aslin
Marc Atkins
Sally Atkins-Burnett
Germine Awad
Catherine Bagwell
Lorraine Bahrick
Jodie Baird
Marian Bakermans-Kranenburg
Lynne Baker-Ward
Marie Balaban
Mayra Bamaca
Mita Banerjee
Lew Bank
Brian Barber
Aron Barbey
David Barner
Andrea Barnes
Melissa Barnett
Andrew Baron
Simon Baron-Cohen
Arthur Baroody
Rachel Barr
Manuel Barrera
Roderick Barron
Dan Bauer
Edith Bavin
Nazli Baydar
Diane Beals
Sarah Beck
Dean Beebe
Harold Bekkering
Debora Bell
Martha Ann Bell
Deborah Belle
Jay Belsky
April Benasich
Joyce Benenson
Aprile Benner
Janette Benson
Sheri Berenbaum
Sarah Berger
Christi Bergin
Daniel Bernstein
Lucy Betts
Ellen Bialystok
Jeremy Biesanz
Ann Bigelow
Rebecca Bigler
Zeynep Biringen
Jeffrey Bisanz
Dorothy Bishop
Angela Black
Maureen Black
Clancy Blair
James Blair
Alysia Blandon
Belinda Blevins-Knabe
Pamela Blewitt
Kerry Bolger
Liz Bonawitz
Julie Booth
Casey Borch
Lera Boroditsky
Heather Bortfeld
Michelle Bosquet

Julie Bowker
W. Thomas Boyce
Ty Boyer
Margaret Bradley
Robert Bradley
Rebecca Brand
Stephen Brand
Holly Branigan
Elizabeth Brannon
J. Bremner
Caitlin Brez
Preston Britner
Pia Britto
Barbara Brizuela
David Brodzinsky
Marion Broome
Laurie Brotman
B. B. Brown
William Brown
Celia Brownell
Christopher Browning
Jacqueline Bruce
Vicki Bruce
Kristen Bub
Martha Buell
William Bukowski
Rebecca Bull
B. Grace Bullock
Raymond Buriel
Virginia Burks-Salzer
Mandi Burnette
Kevin Bush
Kristin Buss
Virginia Buysse
Krista Byers-Heinlein
Brian Byrne
Natasha Cabrera
Josep Call
Tara Callaghan
Maureen Callanan
Lindsey Cameron
Nicole Campione-Barr
Deborah Capaldi
Robert Capraro
Marketa Caravolas
Noel Card
Richard Carlson
Jeremy Carpendale
Malinda Carpenter
Alice Carter
Jocelyn Carter
Leslie Carver
Marianella Casasola
Rita Casey
Larry Cashion
Jude Cassidy
Margaret Caughy
Craig Chambers
Kyle Chambers
Charissa Cheah
Xinyin Chen
Nina Chien
Jane Childers
Yoonsun Choi
Michelle Chouinard
Toon Cillessen
Andrei Cimpian
Eve Clark
Steven Clark
Laura Claxton
Susan Cohen Levine
Jeffrey Cohn
Jeffrey Coldren
Rebekah Coley
Malinda Colwell
Arin Connell
Carol Connor
Joan Cook
Thomas Cook
Michele Cooley-Strickland
Carey Cooper
Shauna Cooper
Michael Corballis
Daniela Corbetta
Sara Cordes
Barry Corenblum
Maricela Correa-Chavez
Roberta Corrigan
Kathleen Corriveau
Robert Corwyn
Sylvana Cote
Mary Courage
Lisa Crockett
Danielle Crosby
William Cross Jr.
David Crystal
Gergely Csibra
Mark Cummings
Sarah Cummins-Sebree
Suzanne Curtin
J. J. Cutuli
Dario Cvencek
Chris Daddis
Marvin Daehler
Ronald Dahl
Philip Dale
James Dannemiller
Judith Danovitch
Nancy Darling
Kelly Davis
Pamela Davis-Kean
Gedeon Deák
Eric Dearing
Kirby Deater-Deckard
Deborah Deckner
Ganie DeHart
Maryse Delaunay-El Allam
Richard De Lisi
Yvon Delville
Jack Demick
Katherine Demuth
Joanne Deocampo
Marc de Rosnay
Jill de Villiers
Lisa Diamond
Lisabeth DiLalla
Janet DiPietro
James Dixon
Jennifer Dobbs-Oates
Kenneth Dodge
Chris Dollaghan
Erik Domellof
Wilberta Donovan
Jason Downer
Roger Downs
Chantelle Dowsett
Eric Dubow
Kevin Dunbar
Yarrow Dunham
Anne Dunlea
Michael Dunne
Julie Dunsmore
Aydin Durgunoglu
Kevin Durkin
Joan Durrant
Matthew Dykas
Warren Eaton
Mary Eberly
Rebecca Eilers
Nancy Eisenberg
Holger Elischberger
Mona El-Sheikh
Susan Engel
Michelle Englund
Rosie Ensor
Cynthia Erdley
Kimberly Espy
David Estell
E. Margaret Evans
Gary Evans
Mary Ann Evans
Jacqueline Fagard

Jeffrey Fagen
Kostas Fanti
Brad Farrant
Teresa Farroni
JoAnn Farver
Christine Fawcett
Allard Feddes
Lisa Feigenson
Heidi Feldman
Christopher Fennell
Christopher Ferguson
Anne Fernald
Charles Fernyhough
Sam Field
Barbara Fiese
Jozsef Fiser
Margaret Fish
Cynthia Fisher
Donna Fisher-Thompson
Connie Flanagan
Alison Fleming
Ross Flom
Mary Ann Foley
Paula Fomby
Erika Forbes
Catherine Forestell
Melissa Rae Fortner
Tim Fosker
Doree Fragaszy
Bridget Franks
Brandy Frazier
Ellen Frede
Jennifer Fredricks
Jill Freiberg
Andrea Frick
Janet Frick
Angela Friederici
Ori Friedman
Sarah Friedman
William Friedman
Alice Frye
Lynn Fuchs
Allison Fuligni
Bruce Fuller
Wyndol Furman
Patricia Ganea
Jane Gaultney
Sybil Geldart
Donna Gelfand
Paul Gendreau
Fred Genesee
Dedre Gentner
Carol George
Gyorgy Gergely
Peter Gerhardstein
Elizabeth Gershoff
Simona Ghetti
John Gibbs
Jeffrey Gilger
Marnie Ginsberg
Marika Ginsburg-Block
Jennifer Glick
Laura Glynn
Tilbe Goksun
Susan Golbeck
Adele Goldberg
Laura Gonnerman
Nancy Gonzales
Gail Goodman
Jacqueline Goodnow
Rachel Gordon
Katharine Graf-Estes
James Green
Patricia Greenfield
Andrea Greenhoot
Marissa Greif
Judith Griffiths
Kevin Grimm
Joan Grusec
Theres Gruter
John Grych
Michaela Gummerum
Megan Gunnar
Guang Guo
Gwen Gustafson
Grant Gutheil
Robert Guttentag
Louise Hainline
Amy Halberstadt
Graeme Halford
Jacob Ham
Constance Hammen
Carol Hammer
Bridget Hamre
Wen-Jui Han
Gordon Harold
Lauren Harris
Paul Harris
Daniel Hart
Etsuko Haryu
Marcus Hasselhorn
John Hattie
Petra Hauf
Dale Hay
Brett Hayes
Harlene Hayne
Jeffrey Heinz
Charles Helwig
Heather Henderson
Elizabeth Hennon
Gaby Hermon
Matthew Hertenstein
Susan Hespos
Gail Heyman
Carri Hill
Melissa Hines
Ernest Hodges
Sara Hodges
Erika Hoff
Julie Hofheimer
Wendy Hoglund
George Holden
George Hollich
Alice Honig
Hyman Hops
Stacey Horn
Derek Houston
Mark Howe
Nina Howe
Carollee Howes
Chi-Tai Huang
Julie Hubbard
Cynthia Hudley
Carla Hudson Kam
L. Rowell Huesmann
Alycia Hund
Carol Huntsinger
Aquiles Iglesias
Kalypso Iordanou
Iheoma Iruka
Miles Irving
Shoji Itakura
Jana Iverson
Fiona Jack
Sandra Jacobson
Sophie Jacques
Laudan Jahromi
Amy Joh
Wendy Johnson
Judith Johnston
Julie Jones
Susan Jones
Nancy Jordan
Sandra Josef-Bishop
Ernest Jouriles
Linda Juang
Femmie Juffer
Laura Justice
Kirsten Kainz
Nina Kaiser
Kathleen Kannass
Grace Kao

Rachel Karchmer-Klein
Annette Karmiloff-Smith
Katherine Karraker
Sheryl Kataoka
Brian Katz
Bruce Kay
Kate Keenan
Frank Keil
Deb Kelemen
Heidi Keller
Peggy Keller
Deborah Kemler Nelson
Margaret Kern
David Kerr
Alan Kersten
Boaz Keysar
Brett Kia-Keating
Maryam Kia-Keating
Lisa Kiang
Ji-Yeon Kim
Sara King
Katherine Kinzler
Katherine Kipp
Paul Klaczynski
David Klahr
Thomas Klee
Matthias Kliegel
George Knight
Becky Kochenderfer-Ladd
Melissa Koenig
Silvia Koller
Agnes Kovacs
Helena Kraemer
Ann Kruger
Jennifer Krull
Patricia Kuhl
Sarah Kulkofsky
Gabriel Kuperminc
Beth Kurtz-Costes
Tamar Kushnir
Gary Ladd
Deborah Laible
Robert Laird
Laura Lakusta
Sharon Lambert
Susan Landry
Daniel Lapsley
Mathieu Le Corre
Amy Learmonth
Joanna Lee
Kang Lee
Esther Leerkes
Cristine Legare
Michelle Leichtman
Patrick Leman
Elizabeth Lemerise
Kathryn Lemery
Liliana Lengua
Richard Lerner
Nonie Lesaux
Leslie Leve
Tama Leventhal
Iris Levin
Sheri Levy
Yonata Levy
Kelly Lewis
Michael Lewis
Peggy Li
Lynn Liben
Oliver Lindhiem
Michael Lindsey
Deborah Linebarger
Miriam Linver
David Liu
Tania Lombrozo
Christopher Lonigan
Michael Lopez
Catherine Lord
Betsy Lozoff
Erika Lunkenheimer
Yuyan Luo
Thomas Lyon
Karlen Lyons-Ruth
Lili Ma
Weiyi Ma
Viola Macchi Cassia
Brian MacWhinney
Katherine Magnuson
Francoise Maheu
Joseph Mahoney
Maria Makrides
Carolina Maldonado-Carreño
Tina Malti
Sarah Mangelsdorf
Virginia Marchman
Stuart Marcovitch
Ellen Markman
Amy Marks
David Marks
Peter Marshall
Carol Martin
Clay Mash
Andy Mashburn
Dominic Massaro
Murray Maybery
Matthew Mayer
Linda Mayes
Ashley Maynard
Keith McBurnett
Teresa McCormack
Lorraine McCune
Jessica McDermott Sales
Nancy McElwain
Heidi McGlothlin
Shirley McGuire
Megan McIlreavy
Thomas McMahon
Nicole McNeil
Douglas Medin
Julie Meeks Gardner
Wim Meeus
Elizabeth Meins
Annika Melinder
Zena Mello
Andrew Meltzoff
Purnima Menon
Paula Menyuk
Sterett Mercer
William Merriman
Carolyn Mervis
Emily Messersmith
Daniel Messinger
Meredith Meyer
George Michel
Amori Mikami
Cindy Miller
Kevin Miller
Karen Milligan
Rashmita Mistry
Christina Mitchell
Kelly Mix
Richard Mizerski
Victoria Molfese
Derek Montgomery
Chris Moore
Michael Morales
Amanda Morris
Anne Morris
Frederick Morrison
J. Morton
Margaret Moulson
Ulrich Mueller
Peter Mundy
Leslie Nabors Olah
Agnes Nairn
Chandan Narayan
Elizabeth Nawrot
Jenae Neiderhiser
Charles Nelson
Ted Nettelbeck
John Newman
Rochelle Newman

Florrie Fei-Yin Ng
Simone Nguyen
Ageliki Nicolopoulou
Mark Nielsen
Adrienne Nishina
Gavin Nobes
Ira Noveck
Larry Nucci
Erika Nurmsoo
Lisa Oakes
Jelena Obradovic
Laurie O'Brien
Thomas O'Connor
Candice Odgers
Yukari Okamoto
Kristina Olson
Kristine Onishi
Mirjam Oosterman
John Opfer
Alexander Ortega
Jamie Ostrov
G. Ouellette
William Overman
Julie Owens
Kerstin Pahl
Robin Panneton
Scott Paris
Julia Parish-Morris
Ty Partridge
Thanujeni Pathman
Meagan Patterson
Barbara Pearson
Lucinda Pease-Alvarez
Kathleen Peets
Anthony Pellegrini
Liz Pellicano
Patricia Pendry
Bruce Pennington
Koraly Perez-Edgar
Amy Perfors
Susan Perlman
Josef Perner
David Perry
Penny Pexman
Jennifer Pfeifer
Alison Pike
Karen Pine
Laura Pittman
Robert Plomin
Seth Pollak
Antonio Polo
Suet-ling Pong
Diane Poulin-Dubois
Thomas Power
Kimberly Powlishta
Sandeep Prasada
Michael Pratt
Kristopher Preacher
Hazel Prelow
Joseph Price
Shannon Pruden
Peter Pufall
Elizabeth Pungello
Jodi Quas
Catherine Raeff
Gnanathusharan Rajendran
Lakshmi Raman
Michael Ramscar
Nan Ratner
C. Raver
Elaine Reese
Stephanie Reich
Ethan Remmel
Jennifer Rennels
Betty Repacholi
Leslie Rescorla
Gary Resnick
David Rettew
J. Steven Reznick
Mijke Rhemtulla
Marjorie Rhodes
Mabel Rice
John Richards
Rebekah Richert
Lindsey Richland
Loretta Rieser-Danner
Tracy Riggins
Kevin Riggs
Giulia Righi
Kristen Ritchey
Rachel Robbins
Kim Roberts
Elizabeth Robinson
Philippe Rochat
Phillip Rodkin
Monica Rodriguez
Claudia Roebers
Robert Roeser
Michael Rohwer
Amanda Rose
Sarah Roseberry
Linda Rose-Krasnor
Karl Rosengren
Hildy Ross
Meredith Rowe
Martin Rowley
Stephanie Rowley
Kelly Rulison
Julie Rusby
Alan Russell
Sandra Saavedra
Abraham Sagi-Schwartz
Roger Saljo
Christina Salmivalli
Kristen Salomon
Elliot Saltzman
Catherine Sandhofer
Marlene Sandstrom
Barbara Sarnecka
Cliff Saron
Megan Saylor
Laura Scaramella
Suzy Scherf
Anne Schlottmann
Stephanie Schmitz
Mark Schmuckler
Michael Schoeny
Christine Schuette
David Schultz
Laura Schulz
Paula Schwanenflugel
Jason Scofield
Lisa Scott
Ronald Seifer
Monique Sénéchal
Maria Sera
Nitya Sethuraman
David Shaenfield
Patrick Shafto
Lilly Shanahan
Timothy Shanahan
Daniel Shaw
Rushen Shi
Rebecca Shiner
Jeanne Shinskey
Elizabeth Shirtcliff
Andrew Shtulman
Anna Shusterman
Kristin Shutts
David Shwalb
Margaret Signorella
Vanessa Simmering
Lynn Singer
Stefanie Sinno
Sylvain Sirois
Lance Slade
Virginia Slaughter
Dan Slobin
Patricia Smiley
Cynthia Smith
Melissa Smith
Lisa Smithers

Keith Smolkowski
Jesse Snedeker
Nancy Snidman
Catherine Snow
David Sobel
Bryan Sokol
Gregg Solomon
Hyun-joo Song
Lulul Song
Jennifer Sootsman Buresh
Christia Spears Brown
Marcia Spetch
Daniel Spieler
Donna Spiker
Tracy Spinrad
Keith Stanovich
Kathy Stansbury
Laurence Steinberg
Diane St-Laurent
Thomas Stoffregen
Mike Stoolmiller
Holly Storkel
Mark Strauss
Tricia Striano
Melissa Sturge-Apple
Thomas Suddendorf
Andrew Supple
Luca Surian
Amy Susman-Stillman
Daniel Swingley
Moin Syed
Marianna Szabo
Lisa Szechter
Joel Talcott
Paula Tallal
Catherine Tamis-LeMonda
Akaysha Tang
Twila Tardif
Lorraine C. Taylor
Marjorie Taylor
Harriet Tenenbaum
Andrew Terranova
Douglas Teti
Ulrich Teucher
David Thomas
Malathi Thothathiri
Peter Titzmann
Michael Tomasello
Bruce Tomblin
Satoshi Tomioka
Maggie Toplak
Nim Tottenham
Dominique Treboux
Rebecca Treiman
Georgene Troseth
Elliot Turiel
Khara Turnbull
David Tzuriel
Wadiya Udell
Tim Urdan
Carmen Valdez
Carlos Valiente
Claire Vallotton
Manfred van Dulmen
Aaron Vaughn
Brian Vaughn
René Veenstra
Alison Ventura
Beth Venzke
Elizabeth Vera
Beatrix Vereijken
Maykel Verkuyten
Elisabetta Visalberghi
Frank Vitaro
Brenda Volling
Virginia Volterra
Claes von Hofsten
Maria von Salisch
Jennifer Vonk
Athena Vouloumanos
Theodore Wachs
Laura Wagner
Ina Wallace
Stephen Want
Liza Ware
Felix Warneken
Heather Warren
Amanda Waterman
Harriet Waters
Norman Watt
Roderick Watts
Niobe Way
Nancy Weinfield
Marsha Weinraub
Deena Weisberg
Brandon Welsh
Marilyn Welsh
Xiaoli Wen
Janet Werker
Donald Wertlieb
Jerry West
Richard West
Paul Whalen
Barbara White
Rhonda White-Johnson
Keith Widaman
Sherri Widen
W. Wilcox
Kirk Williams
Mark Wilson
Melvin Wilson
Ellen Winner
David Witherington
Melissa Witkow
Dustin Wood
Elizabeth Woodburn
Amanda Woodward
John Worobey
Yiyuan Xu
Christine Yeh
Wei-Jun Yeung
Tiffany Yip
Hirokazu Yoshikawa
Sukkyung You
Lise Youngblade
Jonathan Zaff
Luis Zayas
Debra Zeifman
Janice Zeman
Qing Zhou
Kathleen Ziol-Guest
Patricia Zukow-Goldring

Child Development, November/December 2010, Volume 81, Number 6, Pages 1917–1924

Index to Volume 81

Abramovitz, Robert *See* Chemtob, Claude M.

Adam, Hubertus *See* Klasen, Fionna

Adi-Japha, Esther; Berberich-Artzi, Jennie; & Libnawi, Afaf. Cognitive Flexibility in Drawings of Bilingual Children, 1356–1366

Adler, Nancy E. *See* Obradović, Jelena

Adolph, Karen E. *See* Berger, Sarah E.

Ager, Alastair; Stark, Lindsay; Akesson, Bree; & Boothby, Neil. Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study, 1271–1286

Ahern, Elizabeth C. *See* Lyon, Thomas D.

Akesson, Bree *See* Ager, Alastair

Ambady, Nalini *See* Pauker, Kristin

Apfelbaum, Evan P. *See* Pauker, Kristin

Arslanagić, Berina *See* Layne, Christopher M.

Astington, Janet W. *See* Filippova, Eva

Bachman, Heather J. *See* El Nokali, Nermeen E.

Baker, Aaron *See* Layne, Christopher M.

Bakermans-Kranenburg, Marian J. *See* Dobrova-Krol, Natasha A.

Bakermans-Kranenburg, Marian J. *See* Fearon, R. Pasco

Bao, Xuehua *See* Xu, Fen

Barbarin, Oscar A. *See* Chien, Nina C.

Bardi, Lara *See* Turati, Chiara

Barrière, Isabelle *See* Legendre, Géraldine

Bassok, Daphna. Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups, 1828–1845

Becker-Blease, Kathryn A.; Turner, Heather A.; & Finkelhor, David. Disasters, Victimization, and Children's Mental Health, 1040–1052

Belsky, Jay *See* Vandell, Deborah L.

Ben-Eliyahu, Adar *See* Birnbaum, Dana

Benjamin Hinnant, J. *See* El-Sheikh, Mona

Berberich-Artzi, Jennie *See* Adi-Japha, Esther

Berger, Dale E. *See* Fernando, Gaithri A.

Berger, Sarah E.; Adolph, Karen E.; & Kavookjian, Alison E. Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task, 1367–1375

Bering, Jesse M. *See* Ingram, Gordon P. D.

Bernier, Annie; Carlson, Stephanie M.; & Whipple, Natasha. From External Regulation to Self-Regulation: Early Parenting Precursors of Young Children's Executive Functioning, 326–339

Bernier, Annie; Carlson, Stephanie M.; Bordeleau, Stéphanie; & Carrier, Julie. Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning, 1739–1752

Best, John R.; & Miller, Patricia H. A. Developmental Perspective on Executive Function, 1641–1660

Betancourt, Theresa S.; Borisova, Ivelina I.; Williams, Timothy P.; Brennan, Robert T.; Whitfield, Theodore H.; de la Soudiere, Marie; Williamson, John; & Gilman, Stephen E. Sierra Leone's Former Child Soldiers: A Follow-Up Study of Psychosocial Adjustment and Community Reintegration, 1077–1095

Bigler, Rebecca S. *See* Weisgram, Erica S.

Birnbaum, Dana; Deeb, Inas; Segall, Gili; Ben-Eliyahu, Adar; & Diesendruck, Gil. The Development of Social Essentialism: The Case of Israeli Children's Inferences About Jews and Arabs, 757–777

Bisanz, Jeffrey *See* LeFevre, Jo-Anne

Björnberg, Marina *See* Leman, Patrick J.

Bolton, Patrick *See* Ronald, Angelica

Boothby, Neil *See* Ager, Alastair

Bordeleau, Stéphanie *See* Bernier, Annie

Borisova, Ivelina I. *See* Betancourt, Theresa S.

Bornstein, Marc H.; & Mash, Clay. Experience-Based and On-Line Categorization of Objects in Early Infancy, 884–897

Bosacki, Sandra Leanne *See* Lagattuta, Kristin Hansen

Boseovski, Janet J. *See* Marcovitch, Stuart

Boyce, W. Thomas *See* Obradović, Jelena

Boyle, Michael *See* Schmidt, Louis A.

Branje, Susan *See* Meeus, Wim

Brennan, Adrianne M. *See* Kronenberg, Mindy E.

Brennan, Robert T. *See* Betancourt, Theresa S.

Brotman, Laurie Miller *See* O'Neal, Colleen R.

Brownell, Celia A. *See* Svetlova, Margarita

Brownell, Celia A.; Nichols, Sara R.; Svetlova, Margarita; Zerwas, Stephanie; & Ramani, Geetha. The Head Bone's Connected to the Neck Bone: When Do Toddlers Represent Their Own Body Topography? 797–810

Bruce, Jacqueline *See* Pears, Katherine C.

Bryant, Donna M. *See* Chien, Nina C.

Bryant, Joan D. *See* Fuchs, Lynn S.

Buckhalt, Joseph A. *See* El-Sheikh, Mona

Burchinal, Margaret *See* Chien, Nina C.

Burchinal, Margaret *See* Pungello, Elizabeth P.

Burchinal, Margaret *See* Vandell, Deborah L.

Bush, Nicole R. *See* Obradović, Jelena

Bushman, Brad J. *See* Thomaes, Sander

Calzada, Esther J. *See* O'Neal, Colleen R.

Ćampara, Nihada *See* Layne, Christopher M.

Campbell, Frances A. *See* Pungello, Elizabeth P.
Campione-Barr, Nicole; & Smetana, Judith G. "Who Said You Could Wear My Sweater?" Adolescent Siblings' Conflicts and Associations With Relationship Quality, 464–471
Carlson, Stephanie M. *See* Bernier, Annie
Carmel, Nurit *See* Diesendruck, Gil
Carpenter, Malinda *See* Vaish, Amrisha
Carrier, Julie *See* Bernier, Annie
Cassidy, Jude *See* Dykas, Matthew J.
Catani, Claudia; Gewirtz, Abigail H.; Wieling, Elizabeth; Schauer, Elisabeth; Elbert, Thomas; & Neuner, Frank. Tsunami, War, and Cumulative Risk in the Lives of Sri Lankan Schoolchildren, 1176–1191
Chang, Lei *See* Gershoff, Elizabeth T.
Chemtob, Claude M.; Nomura, Yoko; Rajendran, Khushmand; Yehuda, Rachel; Schwartz, Deena; & Abramovitz, Robert. Impact of Maternal Posttraumatic Stress Disorder and Depression Following Exposure to the September 11 Attacks on Preschool Children's Behavior, 1129–1141
Chen, Yu-ping; Keen, Rachel; Rosander, Kerstin; & von Hofsten, Claes Movement Planning Reflects Skill Level and Age Changes in Toddlers, 1846–1858
Chien, Nina C.; Howes, Carollee; Burchinal, Margaret; Pianta, Robert C.; Ritchie, Sharon; Bryant, Donna M.; Clifford, Richard M.; Early, Diane M.; & Barbarin, Oscar A. Children's Classroom Engagement and School Readiness Gains in Prekindergarten, 1534–1549
Cicchetti, Dante; Rogosch, Fred A.; Gunnar, Megan R.; & Toth, Sheree L. The Differential Impacts of Early Physical and Sexual Abuse and Internalizing Problems on Daytime Cortisol Rhythm in School-Aged Children, 252–269
Cicchetti, Dante; Rogosch, Fred A.; Howe, Mark L.; & Toth, Sheree L. The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance, 1504–1519
Clifford, Richard M. *See* Chien, Nina C.
Coley, Rebekah L. *See* Votruba-Drzal, Elizabeth
Compton, Donald L. *See* Fuchs, Lynn S.
Conger, Rand *See* Leve, Leslie D.
Correa-Chávez, Maricela *See* Silva, Katie G.
Corriveau, Kathleen H. *See* Sobel, David M.
Costigan, Kathleen A. *See* DiPietro, Janet A.
Cristi-Vargas, Renate *See* Rafetseder, Eva
Crosnoe, Robert; Leventhal, Tama; Wirth, R. J.; Pierce, Kim M.; Pianta, Robert C.; & NICHD Early Child Care Research Network, Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood, 972–987
Crouter, Ann C. *See* Wray-Lake, Laura

Daniels, Judith *See* Klasen, Fionna
Đapo, Nermin *See* Layne, Christopher M.
Davidov, Maayan *See* Grusec, Joan E.
Davies, Patrick T. *See* Sturge-Apple, Melissa L.
Davis, Elysia P.; & Sandman, Curt A. The Timing of Prenatal Exposure to Maternal Cortisol and Psychosocial Stress Is Associated With Human Infant Cognitive Development, 131–148
Deater-Deckard, Kirby *See* Gershoff, Elizabeth T.
de Castro, Bram O. *See* Thomaes, Sander
Deeb, Inas *See* Birnbaum, Dana
DeGarmo, David Scott. Coercive and Prosocial Fathering, Antisocial Personality, and Growth in Children's Postdivorce Noncompliance, 503–516
de la Soudiere, Marie *See* Betancourt, Theresa S.
Diesendruck, Gil *See* Birnbaum, Dana
Diesendruck, Gil; Carmel, Nurit; & Markson, Lori Children's Sensitivity to the Conventionality of Sources, 652–668
Dijkstra, Jan K. *See* Veenstra, René
DiPietro, Janet A.; Kivlighan, Katie T.; Costigan, Kathleen A.; Rubin, Suzanne E.; Shiffler, Dorothy E.; Henderson, Janice L.; & Pillion, Joseph P. Prenatal Antecedents of Newborn Neurological Maturation, 115–130
Doan, Stacey N. *See* Fuller-Rowell, Thomas E.
Doan, Stacey N.; & Wang, Qi. Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children, 1490–1503
Dobrova-Krol, Natasha A.; van IJzendoorn, Marinus H.; Bakermans-Kranenburg, Marian J.; & Juffer, Femmie. Effects of Perinatal HIV Infection and Early Institutional Rearing on Physical and Cognitive Development of Children in Ukraine, 237–251
Dodge, Kenneth A. *See* Gershoff, Elizabeth T.
D'Onofrio, Brian M.; Singh, Amber L.; Iliadou, Anastasia; Lambe, Mats; Hultman, Christina M.; Neiderhiser, Jenae M.; Långström, Niklas; & Lichtenstein, Paul. A Quasi-Experimental Study of Maternal Smoking During Pregnancy and Offspring Academic Achievement, 80–100
Duncan, Greg J.; Ziol-Guest, Kathleen M.; & Kalil, Ariel Early-Childhood Poverty and Adult Attainment, Behavior, and Health, 306–325
Dunn, Judy. Commentary and Challenges to Grusec and Davidov's Domain-Specific Approach, 710–714
Durakovic'-Belko, Elvira *See* Layne, Christopher M.
Dworzynski, Katharina *See* Ronald, Angelica
Dykas, Matthew J.; Woodhouse, Susan S.; Ehrlich, Katherine B.; & Cassidy, Jude. Do Adolescents and Parents Reconstruct Memories About Their Conflict as a Function of Adolescent Attachment? 1445–1459

Early, Diane M. *See* Chien, Nina C.
Egozi, Gila *See* Tzuriel, David
Ehrlich, Katherine B. *See* Dykas, Matthew J.
Eigsti, Inge-Marie *See* Helt, Molly S.

Elbert, Thomas *See* Catani, Claudia

El Nokali, Nermeen E.; Bachman, Heather J.; & Votruba-Drzal, Elizabeth. Parent Involvement and Children's Academic and Social Development in Elementary School, 988–1005

El-Sheikh, Mona; Kelly, Ryan J.; Buckhalt, Joseph A.; & Benjamin Hinnant, J. Children's Sleep and Adjustment Over Time: The Role of Socioeconomic Context, 870–883

Erickson, Jane E.; Keil, Frank C.; & Lockhart, Kristi L. Sensing the Coherence of Biology in Contrast to Psychology: Young Children's Use of Causal Relations to Distinguish Two Foundational Domains, 390–409

Fast, Lisa *See* LeFevre, Jo-Anne

Fearon, R. Pasco; Bakermans-Kranenburg, Marian J.; van Ijzendoorn, Marinus H.; Lapsley, Anne-Marie & Roisman, Glenn I. The Significance of Insecure Attachment and Disorganization in the Development of Children's Externalizing Behavior: A Meta-Analytic Study, 435–456

Fein, Deborah A. *See* Helt, Molly S.

Fennell, Christopher T.; & Waxman, Sandra R. What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning, 1376–1383

Fernando, Gaithri A.; Miller, Kenneth E.; & Berger, Dale E. Growing Pains: The Impact of Disaster-Related and Daily Stressors on the Psychological and Psychosocial Functioning of Youth in Sri Lanka, 1192–1210

Ferry, Alissa L.; Hespos, Susan J.; & Waxman, Sandra R. Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones 472–479

Filippova, Eva; & Astington, Janet W. Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony, 913–928

Finkelhor, David *See* Becker-Blease, Kathryn A.

Fisher, Philip A. *See* Pears, Katherine C.

Flinn, Mark V. *See* Vigil, Jacob M.

Foster, E. Michael; & Watkins, Stephanie The Value of Reanalysis: TV Viewing and Attention Problems, 368–375

Fox, Nathan A. *See* Smyke, Anna T.

Fox, Nathan A.; & Rutter, Michael. Introduction to the Special Section on The Effects of Early Experience on Development, 23–27

Fox, Sharon E.; Levitt, Pat; & Nelson III, Charles A. How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture, 28–40

Frenn, Kristin A. *See* Pollak, Seth D.

Fu, Genyue *See* Xu, Fen

Fuchs, Douglas *See* Fuchs, Lynn S.

Fuchs, Lynn S.; Geary, David C.; Compton, Donald L.; Fuchs, Douglas; Hamlett, Carol L.; & Bryant, Joan D. The Contributions of Numerosity and Domain-General Abilities to School Readiness, 1520–1533

Fuller-Rowell, Thomas E.; & Doan, Stacey N. The Social Costs of Academic Success Across Ethnic Groups, 1696–1713

Ganea, Patricia A.; & Harris, Paul L. Not Doing What You Are Told: Early Perseverative Errors in Updating Mental Representations via Language, 457–463

Ge, Xiaojia *See* Leve, Leslie D.

Geary, David C. *See* Fuchs, Lynn S.

Geary, David C. *See* Vigil, Jacob M.

Gelman, Susan A. *See* Legare, Cristine H.

Gelman, Susan A. Modules, Theories, or Islands of Expertise? Domain Specificity in Socialization, 715–719

Gershoff, Elizabeth T.; Grogan-Kaylor, Andrew; Lansford, Jennifer E.; Chang, Lei; Zelli, Arnaldo; Deater-Deckard, Kirby; & Dodge, Kenneth A. Parent Discipline Practices in an International Sample: Associations With Child Behaviors and Moderation by Perceived Normativeness, 487–502

Gershoff, Elizabeth T.; Lawrence Aber, J.; Ware, Angelica; & Kotler, Jennifer A. Exposure to 9/11 Among Youth and Their Mothers in New York City: Enduring Associations With Mental Health and Sociopolitical Attitudes, 1142–1160

Gewirtz, Abigail H. *See* Catani, Claudia

Gieling, Maike; Thijs, Jochem; & Verkuyten, Maykel. Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective, 1384–1399

Gillespie, Rebecca *See* Wagner, Laura

Gilman, Stephen E. *See* Betancourt, Theresa S.

Gil-Rivas, Virginia *See* Kilmer, Ryan P.

Giorgio, Elisa D. *See* Turati, Chiara

Goldin-Meadow, Susan *See* Sauer, Eve

Gouley, Kathleen Kiely *See* O'Neal, Colleen R.

Goyet, Louise *See* Legendre, Géraldine

Granger, Douglas A. *See* Vigil, Jacob M.

Gredebäck, Gustaf *See* Kochukhova, Olga

Greene-Havas, Maia *See* Wagner, Laura

Grogan-Kaylor, Andrew *See* Gershoff, Elizabeth T.

Grosse, Scott D. Late-Treated Phenylketonuria and Partial Reversibility of Intellectual Impairment, 200–211

Grosser, Tilo *See* Price, Thomas S.

Grusec, Joan E.; & Davidov, Maayan, Integrating Different Perspectives on Socialization Theory and Research: A Domain-Specific Approach, 687–709

Gunnar, Megan R. *See* Cicchetti, Dante

Gunnar, Megan R. *See* Pollak, Seth D.

Gunnar, Megan R.; Kryzer, Erin; Van Ryzin, Mark J.; & Phillips, Deborah A. The Rise in Cortisol in Family Day Care: Associations With Aspects of Care Quality, Child Behavior, and Child Sex, 851–869

Guthrie, Donald *See* Smyke, Anna T.

Hamlett, Carol L. *See* Fuchs, Lynn S.

Hansel, Tonya C. *See* Kronenberg, Mindy E.

Happé, Francesca *See* Ronald, Angelica

Harris, Paul L. *See* Ganea, Patricia A.

Hasselhorn, Marcus *See* Lehmann, Martin

Hauser, Marc D. *See* Vouloumanos, Athena

Hay, Dale F.; Pawlby, Susan; Waters, Cerith S.; Perra, Oliver; & Sharp, Deborah. Mothers' Antenatal Depression and Their Children's Antisocial Outcomes, 149–165
Helt, Molly S.; Eigsti, Inge-Marie; Snyder, Peter J.; & Fein, Deborah A. Contagious Yawning in Autistic and Typical Development, 1620–1631
Henderson, Janice L. *See* DiPietro, Janet A.
Hespos, Susan J. *See* Ferry, Alissa L.
Hill, Laura G. *See* Werner, Nicole E.
Hilliard, Lacey J.; & Liben, Lynn S. Differing Levels of Gender Salience in Preschool Classrooms: Effects on Children's Gender Attitudes and Intergroup Bias, 1787–1798
Hofferth, Sandra L. Home Media and Children's Achievement and Behavior, 1598–1619
Howe, Mark L. *See* Cicchetti, Dante
Howes, Carollee *See* Chien, Nina C.
Hoyer, Catrin *See* Klasen, Fionna
Huang, Keng-Yen *See* O'Neal, Colleen R.
Hultman, Christina M. *See* D'Onofrio, Brian M.

Iliadou, Anastasia *See* D'Onofrio, Brian M.
Ingram, Gordon P. D.; & Bering, Jesse M. Children's Tattling: The Reporting of Everyday Norm Violations in Preschool Settings, 945–957
Isakson, Brian *See* Layne, Christopher M.
Itakura, Shoji *See* Okanda, Mako

Jaffee, Sara R. *See* Price, Thomas S.
Johnson, Kerri L.; Lurye, Leah E.; & Tassinary, Louis G. Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues, 1346–1355
Jonkman, Lisa M. *See* Vlamings, Petra H. J. M.
Juffer, Femmie *See* Dobrova-Krol, Natasha A.

Kainz, Kirsten *See* Pungello, Elizabeth P.
Kalil, Ariel *See* Duncan, Greg J.
Kamawar, Deepthi *See* LeFevre, Jo-Anne
Kamboukos, Dimitra *See* O'Neal, Colleen R.
Kane, Michael J. *See* Marcovitch, Stuart
Kärtner, Joscha; Keller, Heidi; & Yovsi, Relindis D. Mother–Infant Interaction During the First 3 Months: The Emergence of Culture-Specific Contingency Patterns, 540–554
Kavookjian, Alison E. *See* Berger, Sarah E.
Keen, Rachel *See* Chen, Yu-ping
Keijsers, Loes *See* Meeus, Wim
Keil, Frank C. *See* Erickson, Jane E.
Keller, Heidi *See* Kärtner, Joscha
Kelly, Ryan J. *See* El-Sheikh, Mona
Kemner, Chantal *See* Vlamings, Petra H. J. M.
Kerr, David C. R. *See* Leve, Leslie D.
Kilmer, Ryan P.; & Gil-Rivas, Virginia. Exploring Posttraumatic Growth in Children Impacted by Hurricane Katrina: Correlates of the Phenomenon and Developmental Considerations, 1211–1227
Kim, Hyoun K. *See* Pears, Katherine C.
Kithakye, Mumbe; Morris, Amanda S.; Terranova, Andrew M.; & Myers, Sonya S. The Kenyan Political Conflict and Children's Adjustment, 1114–1128
Kivlighan, Katie T. *See* DiPietro, Janet A.
Klasen, Fionna; Oettingen, Gabriele; Daniels, Judith; Post, Manuela; Hoyer, Catrin; & Adam, Hubertus. Posttraumatic Resilience in Former Ugandan Child Soldiers, 1096–1113
Knapp, Robin J. *See* Marcovitch, Stuart
Kochukhova, Olga; & Gredebäck, Gustaf. Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons, 1729–1738
Kotler, Jennifer A. *See* Gershoff, Elizabeth T.
Kronenberg, Mindy E.; Hansel, Tonya C.; Brennan, Adrianne M.; Osofsky, Howard J.; Osofsky, Joy D.; & Lawrason, Beverly. Children of Katrina: Lessons Learned About Postdisorder Symptoms and Recovery Patterns, 1241–1259
Kryzer, Erin *See* Gunnar, Megan R.

Lagattuta, Kristin Hansen; Nucci, Larry; & Bosacki, Sandra Leanne Bridging Theory of Mind and the Personal Domain: Children's Reasoning About Resistance to Parental Control, 616–635
Lambe, Mats *See* D'Onofrio, Brian M.
Lane, Jonathan D.; Wellman, Henry M.; & Margaret Evans, E. Children's Understanding of Ordinary and Extraordinary Minds, 1475–1489
Långström, Niklas *See* D'Onofrio, Brian M.
Lansford, Jennifer E. *See* Gershoff, Elizabeth T.
Lapsley, Anne-Marie, *See* Fearon, R. Pasco
Lawrason, Beverly *See* Kronenberg, Mindy E.
Lawrence Aber, J. *See* Gershoff, Elizabeth T.
Layne, Christopher M.; Olsen, Joseph A.; Baker, Aaron; Legerski, John-Paul; Isakson, Brian; Pašalić, Alma; Duraković-Belko, Elvira; Đapo, Nermin; Ćampara, Nihada; Arslanagić, Berina; Saltzman, William R.; & Pynoos, Robert S. Unpacking Trauma Exposure Risk Factors and Differential Pathways of Influence: Predicting Postwar Mental Distress in Bosnian Adolescents, 1053–1076
Lecce, Serena; Zocchi, Silvia; Pagnin, Adriano; Palladino, Paola; & Taumoepeau, Mele. Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading, 1876–1893
Lee, Kang *See* Xu, Fen
LeFevre, Jo-Anne; Fast, Lisa; Skwarchuk, Sheri-Lynn; Smith-Chant, Brenda L.; Bisanz,, Jeffrey; Kamawar, Deepthi; & Penner-Wilger, Marcie. Pathways to Mathematics: Longitudinal Predictors of Performance, 1753–1767

Legare, Cristine H.; Gelman, Susan A.; & Wellman, Henry M. Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning, 929–944

Legendre, Géraldine; Barrière, Isabelle; Goyet, Louise; & Nazzi, Thierry. Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children, 1859–1875

Legerski, John-Paul *See* Layne, Christopher M.

Lehmann, Martin; & Hasselhorn, Marcus. The Dynamics of Free Recall and Their Relation to Rehearsal Between 8 and 10 Years of Age, 1006–1020

Leman, Patrick J.; & Björnberg, Marina. Conversation, Development and Gender: A Study of Changes in Children's Concepts of Punishment, 958–971

Leve, Leslie D.; Kerr, David C. R.; Shaw, Daniel; Ge, Xiaojia; Neiderhiser, Jenae M.; Scaramella, Laura V.; Reid, John B.; Conger, Rand; & Reiss, David. Infant Pathways to Externalizing Behavior: Evidence of Genotype x Environment Interaction, 340–356

Leventhal, Tama *See* Crosnoe, Robert

Levine, Susan C. *See* Sauer, Eve

Levitt, Pat *See* Fox, Sharon E.

Li, Yexin Jessica *See* Wang, Qi

Liben, Lynn S. *See* Hilliard, Lacey J.

Liben, Lynn S. *See* Weisgram, Erica S.

Libnawi, Afaf *See* Adi-Japha, Esther

Lichtenstein, Paul *See* D'Onofrio, Brian M.

Li-Grining, Christine P. *See* Votruba-Drzal, Elizabeth

Lindenberg, Siegwart *See* Veenstra, René

Lindsay Chase-Lansdale, P. *See* Votruba-Drzal, Elizabeth

Linebarger, Deborah L.; & Piotrowski, Jessica T. Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning, 1582–1597

Lockhart, Kristi L. *See* Erickson, Jane E.

Loman, Michelle M. *See* Pollak, Seth D.

Low, Jason Preschoolers' Implicit and Explicit False-Belief Understanding: Relations With Complex Syntactical Mastery, 597–615

Lurye, Leah E. *See* Johnson, Kerri L.

Lyon, Thomas D.; Ahern, Elizabeth C.; Malloy, Lindsay C.; & Quas, Jodi A. Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Identity, 1714–1728

Maldonado-Carreño, Carolina *See* Votruba-Drzal, Elizabeth

Malloy, Lindsay C. *See* Lyon, Thomas D.

Marcovitch, Stuart; Boseovski, Janet J.; Knapp, Robin J.; & Kane, Michael J. Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children, 1687–1695

Margaret Evans, E. *See* Lane, Jonathan D.

Mark Cummings, E. *See* Sturge-Apple, Melissa L.

Markson, Lori *See* Diesendruck, Gil

Martin, Alia *See* Vouloumanos, Athena

Mash, Clay *See* Bornstein, Marc H.

Masten, Ann S.; & Osofsky, Joy D. Disasters and Their Impact on Child Development: Introduction to the Special Section, 1029–1039

McHale, Susan M. *See* Wray-Lake, Laura

Meaney, Michael J. Epigenetics and the Biological Definition of Gene × Environment Interactions, 41–79

Meeus, Wim; van de Schoot, Rens; Keijsers, Loes; Schwartz, Seth J.; & Branje, Susan. On the Progression and Stability of Adolescent Identity Formation: A Five-Wave Longitudinal Study in Early-to-Middle and Middle-to-Late Adolescence, 1565–1581

Miller, Kenneth E. *See* Fernando, Gaithri A.

Miller, Patricia H. *See* Best, John R.

Miskovic, Vladimir *See* Schmidt, Louis A.

Morris, Amanda S. *See* Kithakye, Mumbe

Munniksma, Anke *See* Veenstra, René

Myers, Sonya S. *See* Kithakye, Mumbe

Nazzi, Thierry *See* Legendre, Géraldine

Neiderhiser, Jenae M. *See* D'Onofrio, Brian M.

Neiderhiser, Jenae M. *See* Leve, Leslie D.

Nelson III, Charles A. *See* Fox, Sharon E.

Nelson, Charles A. *See* Pollak, Seth D.

Nelson, Charles A. *See* Smyke, Anna T.

Neuner, Frank *See* Catani, Claudia

NICHD Early Child Care Research Network *See* Crosnoe, Robert

NICHD Early Child Care Research Network *See* Vandell, Deborah L.

Nichols, Sara R. *See* Brownell, Celia A.

Nichols, Sara R. *See* Svetlova, Margarita

Nomura, Yoko *See* Chemtob, Claude M.

Nucci, Larry *See* Lagattuta, Kristin Hansen

Obradović, Jelena; Bush, Nicole R.; Stamperdahl, Juliet; Adler, Nancy E.; & Boyce, W. Thomas. Biological Sensitivity to Context: The Interactive Effects of Stress Reactivity and Family Adversity on Socioemotional Behavior and School Readiness, 270–289

Oettingen, Gabriele *See* Klasen, Fionna

Okanda, Mako; & Itakura, Shoji. When Do Children Exhibit a "Yes" Bias? 568–580

Olsen, Joseph A. *See* Layne, Christopher M.

Oncu, Elif C.; & Wise, Aysegul M. The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories, 1161–1175

O'Neal, Colleen R.; Brotman, Laurie Muller; Huang, Keng-Yen; Gouley, Kathleen Kiely; Kamboukos, Dimitra; Calzada, Esther J.; & Pine, Daniel S. Understanding Relations Among Early Family Environment,

Cortisol Response, and Child Aggression via a Prevention Experiment, 290–305
Opfer, John E. *See* Thompson, Clarissa A.
Osofsky, Howard J. *See* Kronenberg, Mindy E.
Osofsky, Joy D. *See* Kronenberg, Mindy E.
Osofsky, Joy D. *See* Masten, Ann S.
Ostrov, Jamie M. Prospective Associations Between Peer Victimization and Aggression, 1670–1677

Pagnin, Adriano *See* Lecce, Serena
Palladino, Paola *See* Lecce, Serena
Pašalić, Alma *See* Layne, Christopher M.
Pauker, Kristin; Ambady, Nalini; & Apfelbaum, Evan P. Race Salience and Essentialist Thinking in Racial Stereotype Development, 1799–1813
Pawlby, Susan *See* Hay, Dale F.
Pears, Katherine C.; Fisher, Philip A.; Bruce, Jacqueline; Kim, Hyoun K.; & Yoerger, Karen. Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement, 1550–1564
Peek, Lori; & Stough, Laura M. Children With Disabilities in the Context of Disaster: A Social Vulnerability Perspective, 1260–1270
Pellicano, Elizabeth. The Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study, 1400–1416
Penner-Wilger, Marcie *See* LeFevre, Jo-Anne
Perner, Josef *See* Rafetseder, Eva
Perra, Oliver *See* Hay, Dale F.
Phillips, Deborah A. *See* Gunnar, Megan R.
Pianta, Robert C. *See* Chien, Nina C.
Pianta, Robert C. *See* Crosnoe, Robert
Pierce, Kim M. *See* Crosnoe, Robert
Pillion, Joseph P. *See* DiPietro, Janet A.
Pine, Daniel S. *See* O'Neal, Colleen R.
Piotrowski, Jessica T. *See* Linebarger, Deborah L.
Plomin, Robert *See* Price, Thomas S.
Plomin, Robert *See* Ronald, Angelica
Pollak, Seth D.; Nelson, Charles A.; Schlaak, Mary F.; Roeber, Barbara J.; Wewerka, Sandi S.; Wiik, Kristen L.; Frenn, Kristin A.; Loman, Michelle M.; & Gunnar, Megan R. Neurodevelopmental Effects of Early Deprivation in Postinstitutionalized Children, 224–236
Poorthuis, Astrid *See* Thomaes, Sander
Post, Manuela *See* Klasen, Fionna
Price, Thomas S.; Grosser, Tilo; Plomin, Robert; & Jaffee, Sara R. Fetal Genotype for the Xenobiotic Metabolizing Enzyme NQO1 Influences Intrauterine Growth Among Infants Whose Mothers Smoked During Pregnancy, 101–114
Pungello, Elizabeth P.; Kainz, Kirsten; Burchinal, Margaret; Wasik, Barbara H.; Sparling, Joseph J.; Ramey, Craig T.; & Campbell, Frances A. Early Educational Intervention, Early Cumulative Risk, and the Early Home Environment as Predictors of Young Adult Outcomes Within a High-Risk Sample, 410–426
Pynoos, Robert S. *See* Layne, Christopher M.

Quas, Jodi A. *See* Lyon, Thomas D.

Rafetseder, Eva; Cristi-Vargas, Renate; & Perner, Josef. Counterfactual Reasoning: Developing a Sense of 'Nearest Possible World', 376–389
Rajendran, Khushmand *See* Chemtob, Claude M.
Ramani, Geetha *See* Brownell, Celia A.
Ramey, Craig T. *See* Pungello, Elizabeth P.
Reid, John B. *See* Leve, Leslie D.
Reijntjes, Albert *See* Thomaes, Sander
Reiss, David *See* Leve, Leslie D.
Ritchie, Sharon *See* Chien, Nina C.
Rodkin, Philip C.; & Roisman, Glenn I. Antecedents and Correlates of the Popular-Aggressive Phenomenon in Elementary School, 837–850
Roeber, Barbara J. *See* Pollak, Seth D.
Rogoff, Barbara *See* Silva, Katie G.
Rogosch, Fred A. *See* Cicchetti, Dante
Roisman, Glenn I. *See* Fearon, R. Pasco
Roisman, Glenn I. *See* Rodkin, Philip C.
Ronald, Angelica; Happé, Francesca; Dworzynski, Katharina; Bolton, Patrick; & Plomin, Robert. Exploring the Relation Between Prenatal and Neonatal Complications and Later Autistic-Like Features in a Representative Community Sample of Twins, 166–182
Rosander, Kerstin *See* Chen, Yu-ping
Rubin, Suzanne E. *See* DiPietro, Janet A.
Rutter, Michael *See* Fox, Nathan A.

Saffran, Jenny R. *See* Sahni, Sarah D.
Sahni, Sarah D.; Seidenberg, Mark S.; & Saffran, Jenny R. Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy, 727–736
Saigal, Saroj *See* Schmidt, Louis A.
Saltzman, William R. *See* Layne, Christopher M.
Sameroff, Arnold A. Unified Theory of Development: A Dialectic Integration of Nature and Nurture, 6–22
Sandman, Curt A. *See* Davis, Elysia P.
Sauer, Eve; Levine, Susan C.; & Goldin-Meadow, Susan. Early Gesture Predicts Language Delay in Children With Pre- or Perinatal Brain Lesions, 528–539
Scaramella, Laura V. *See* Leve, Leslie D.
Schauer, Elisabeth *See* Catani, Claudia
Schlaak, Mary F. *See* Pollak, Seth D.
Schmidt, Louis A.; Miskovic, Vladimir; Boyle, Michael; & Saigal, Saroj. Frontal Electroencephalogram Asymmetry, Salivary Cortisol, and Internalizing Behavior Problems in Young Adults Who Were Born at Extremely Low Birth Weight, 183–199
Schwartz, Deena *See* Chemtob, Claude M.
Schwartz, Seth J. *See* Meeus, Wim
Seaton, Eleanor K. *See* Yip, Tiffany
Segall, Gili *See* Birnbaum, Dana

Seidenberg, Mark S. *See* Sahni, Sarah D.
Sellers, Robert M. *See* Yip, Tiffany
Shao, Yi *See* Wang, Qi
Sharp, Deborah *See* Hay, Dale F.
Shaw, Daniel *See* Leve, Leslie D.
Shiffler, Dorothy E. *See* DiPietro, Janet A.
Shonkoff, Jack P. Building a New Biodevelopmental Framework to Guide the Future of Early Childhood Policy, 357–367
Silva, Katie G.; Correa-Chávez, Maricela; & Rogoff, Barbara. Mexican-Heritage Children's Attention and Learning From Interactions Directed to Others, 898–912
Simion, Francesca *See* Turati, Chiara
Singh, Amber L. *See* D'Onofrio, Brian M.
Skwarchuk, Sheri-Lynn *See* LeFevre, Jo-Anne
Smetana, Judith G. *See* Campione-Barr, Nicole
Smith-Chant, Brenda L. *See* LeFevre, Jo-Anne
Smyke, Anna T.; Zeanah, Charles H.; Fox, Nathan A.; Nelson, Charles A.; & Guthrie, Donald. Placement in Foster Care Enhances Quality of Attachment Among Young Institutionalized Children, 212–223
Snyder, Peter J. *See* Helt, Molly S.
Sobel, David M.; & Corriveau, Kathleen H. Children Monitor Individuals' Expertise for Word Learning, 669–679
Sparling, Joseph J. *See* Pungello, Elizabeth P.
Stamperdahl, Juliet *See* Obradović, Jelena
Stark, Lindsay *See* Ager, Alastair
Steinberg, Laurence *See* Vandell, Deborah L.
Stough, Laura M. *See* Peek, Lori
Sturge-Apple, Melissa L.; Davies, Patrick T.; & Mark Cummings, E. Typologies of Family Functioning and Children's Adjustment During the Early School Years, 1320–1335
Svetlova, Margarita *See* Brownell, Celia A.
Svetlova, Margarita; Nichols, Sara R.; & Brownell, Celia A. Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping, 1814–1827

Talwar, Victoria *See* Xu, Fen
Tassinary, Louis G. *See* Johnson, Kerri L.
Taumoepeau, Mele *See* Lecce, Serena
Telch, Michael J. *See* Thomaes, Sander
Terranova, Andrew M. *See* Kithakye, Mumbe
Thiessen, Erik D.; & Yee, Meagan N. Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations, 1287–1303
Thijs, Jochem *See* Gieling, Maike
Thomaes, Sander; Reijntjes, Albert; de Castro, Bram O.; Bushman, Brad J.; Poorthuis, Astrid; & Telch, Michael J. I Like Me If You Like Me: On the Interpersonal Modulation and Regulation of Preadolescents' State Self-Esteem, 811–825
Thompson, Clarissa A.; & Opfer, John E. How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition, 1768–1786
Tomasello, Michael *See* Vaish, Amrisha
Toth, Sheree L. *See* Cicchetti, Dante
Turati, Chiara; Giorgio, Elisa D.; Bardi, Lara; & Simion, Francesca. Holistic Face Processing in Newborns, 3-Month-Old Infants, and Adults: Evidence From the Composite Face Effect, 1894–1905
Turiel, Elliot. Domain Specificity in Social Interactions, Social Thought, and Social Development, 720–726
Turner, Heather A. *See* Becker-Blease, Kathryn A.
Tzuriel, David; & Egozi, Gila. Gender Differences in Spatial Ability of Young Children: The Effects of Training and Processing Strategies, 1417–1430

Vaish, Amrisha; Carpenter, Malinda; & Tomasello, Michael. Young Children Selectively Avoid Helping People With Harmful Intentions, 1661–1669
van de Schoot, Rens *See* Meeus, Wim
van IJzendoorn, Marinus H. *See* Dobrova-Krol, Natasha A.
van Ijzendoorn, Marinus H. *See* Fearon, R. Pasco
Van Ryzin, Mark J. *See* Gunnar, Megan R.
Vandell, Deborah L.; Belsky, Jay; Burchinal, Margaret; Steinberg, Laurence; Vandergrift, Nathan; & NICHD Early Child Care Research Network, Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development, 737–756
Vandergrift, Nathan *See* Vandell, Deborah L.
Veenstra, René; Lindenberg, Siegwart; Munniksma, Anke; & Dijkstra, Jan K. The Complex Relation Between Bullying, Victimization, Acceptance, and Rejection: Giving Special Attention to Status, Affection, and Sex Differences, 480–486
Verkuyten, Maykel *See* Gieling, Maike
Vigil, Jacob M.; Geary, David C.; Granger, Douglas A.; & Flinn, Mark V. Sex Differences in Salivary Cortisol, Alpha-Amylase, and Psychological Functioning Following Hurricane Katrina, 1228–1240
Vlamings, Petra H. J. M.; Jonkman, Lisa M.; & Kemner, Chantal. An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children, 1304–1319
von Hofsten, Claes *See* Chen, Yu-ping
Votruba-Drzal, Elizabeth *See* El Nokali, Nermeen E.
Votruba-Drzal, Elizabeth; Coley, Rebekah L.; Maldonado-Carreño, Carolina; Li-Grining, Christine P.; & Lindsay Chase-Lansdale, P. Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood, 1460–1474
Vouloumanos, Athena; Hauser, Marc D.; Werker, Janet F.; & Martin, Alia. The Tuning of Human Neonates' Preference for Speech, 517–527

Wagner, Laura; Greene-Havas, Maia; & Gillespie, Rebecca. Development in Children's Comprehension of Linguistic Register, 1678–1686
Wang, Qi *See* Doan, Stacey N.
Wang, Qi; Shao, Yi; & Li, Yexin J. "My Way or Mom's Way?" The Bilingual and Bicultural Self in Hong Kong Chinese Children and Adolescents, 555–567
Ware, Angelica *See* Gershoff, Elizabeth T.
Wasik, Barbara H. *See* Pungello, Elizabeth P.
Waters, Cerith S. *See* Hay, Dale F.
Watkins, Stephanie *See* Michael Foster, E.
Waxman, Sandra R. *See* Fennell, Christopher T.
Waxman, Sandra R. *See* Ferry, Alissa L.
Weisgram, Erica S.; Bigler, Rebecca S.; & Liben, Lynn S. Gender, Values, and Occupational Interests Among Children, Adolescents, and Adults, 778–796
Wellman, Henry M. *See* Lane, Jonathan D.
Wellman, Henry M. *See* Legare, Cristine H.
Werker, Janet F. *See* Vouloumanos, Athena
Werner, Nicole E.; & Hill, Laura G. Individual and Peer Group Normative Beliefs About Relational Aggression, 826–836
Wewerka, Sandi S. *See* Pollak, Seth D.
Whipple, Natasha *See* Bernier, Annie
Whitfield, Theodore H. *See* Betancourt, Theresa S.
Wieling, Elizabeth *See* Catani, Claudia
Wiik, Kristen L. *See* Pollak, Seth D.
Williams, Timothy P. *See* Betancourt, Theresa S.
Williamson, John *See* Betancourt, Theresa S.
Wirth, R. J. *See* Crosnoe, Robert
Wise, Aysegul M. *See* Oncu, Elif C.
Woodhouse, Susan S. *See* Dykas, Matthew J.
Wray-Lake, Laura; Crouter, Ann C.; & McHale, Susan M. Developmental Patterns in Decision-Making Autonomy Across Middle Childhood and Adolescence: European American Parents' Perspectives, 636–651

Xu, Fen; Bao, Xuehua; Fu, Genyue; Talwar, Victoria; & Lee, Kang Lying and Truth-Telling in Children: From Concept to Action, 581–596

Yee, Meagan N. *See* Thiessen, Erik D.
Yehuda, Rachel *See* Chemtob, Claude M.
Yip, Tiffany; Seaton, Eleanor K.; & Sellers, Robert M. Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents, 1431–1444
Yoerger, Karen *See* Pears, Katherine C.
Yovsi, Relindis D. *See* Kärtner, Joscha

Zeanah, Charles H. *See* Smyke, Anna T.
Zelli, Arnaldo *See* Gershoff, Elizabeth T.
Zerwas, Stephanie *See* Brownell, Celia A.
Ziol-Guest, Kathleen M. *See* Duncan, Greg J.
Zocchi, Silvia *See* Lecce, Serena

UNIVERSITY OF ALABAMA AT BIRMINGHAM: The UAB Department of Psychology is seeking nominations and applications for a new **Endowed Chair in Developmental Psychology**. This is a tenured position at the Associate or Full Professor level. The candidate should be a psychologist with expertise in child or adolescent research and a strong record of scholarship and extramural funding in lifespan development, child clinical or pediatric psychology. The successful candidate for this position will be expected to lead an active and externally funded program of research, mentor graduate and undergraduate students, and will have the opportunity to teach in his/her area of expertise.

The Psychology Department is a vibrant and well regarded department on campus and supports three doctoral programs (Lifespan Developmental, Medical Clinical, and Behavioral Neuroscience). The Department has a long history of obtaining extramural research funding and has excellent research facilities. In addition, as part of the University's emphasis on multidisciplinary centers, the Psychology Department encourages/supports collaboration with various departments and centers within the university including the Civitan International Research Center, the Civitan/Sparks Clinics, The Children's Health System, The Comprehensive Cancer Center, the Center for Neuroscience, The McKnight Brain Institute, The Center for Community Health, The Center for Translational Science, The Center for Aging, and the School of Medicine. Individuals interested in participating, promoting, and advancing a collaborative and multidisciplinary environment at UAB and in the community are encouraged to apply.

The Psychology Department enjoys the benefits of UAB's standing as a national center for biomedical health research and education. Metro Birmingham (population 1.2 million) is an ethnically diverse and cosmopolitan setting with a rich array of cultural institutions and a high quality of life.

Applicants should send a letter of interest, CV, representative reprints, and the names of three professional references to: **Dr. Fred Biasini, Chair, Developmental Psychology Search Committee, Department of Psychology, University of Alabama at Birmingham, 1530 3rd Avenue South, Birmingham, AL 35294-1170. For questions about the positions, please contact Dr. Biasini (fbiasini@uab.edu) or Dr. Karlene Ball (kball@uab.edu), University Professor and Chair, Department of Psychology.** Screening of applications will begin in Fall 2010 and continue until the position is filled. Starting date is negotiable.

UAB is dedicated to broadening the diversity of its faculty, staff, and students. We take pride in our exceptionally diverse and multicultural student body. UAB is an Equal Opportunity/Affirmative Action employer. Women and minority candidates are encouraged to apply.

UAB THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM

Discover *Psychology*
with Wiley-Blackwell

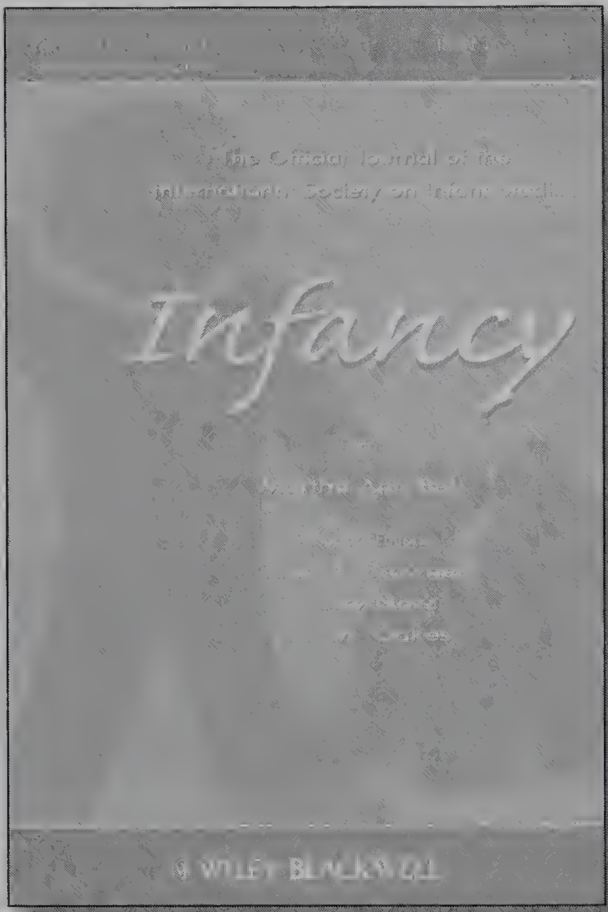
Wiley-Blackwell's esteemed list of psychology books and journals provides the highest quality scholarship for faculty, students, researchers, and practitioners. Subjects in which Wiley-Blackwell publishes include:

- Applied Psychology
- Clinical Psychology
- Cognitive Psychology
- Developmental Psychology
- Educational & School Psychology
- Family Therapy
- Industrial & Organizational Psychology
- Physiological Psychology
- Psychotherapy & Counseling
- Social Psychology

Find all that Wiley-Blackwell has to offer in your field (including links to free content) at
wileyonlinelibrary.com/subject/psychology

WILEY-
BLACKWELL
The Place to Publish

Wiley-Blackwell's industry-leading books and journals publishing program, advanced online delivery platform, high-quality and cost-effective production values, extensive global marketing and sales reach, and state-of-the-art services for societies, editors and authors make Wiley-Blackwell the place to publish across all disciplines.



Infancy

Edited by
MARTHA ANN BELL

Published on behalf of the International Society on Infant Studies

Infancy, the official journal of the International Society on Infant Studies, emphasizes the highest quality original research on normal and aberrant infant development during the first two years. Both human and animal research are included. In addition to regular length research articles and brief reports (3000-word maximum), the journal includes solicited target articles along with a series of commentaries; debates, in which different theoretical positions are presented along with a series of commentaries; and thematic collections, a group of three to five reports or summaries of research on the same issue, conducted independently at different laboratories, with invited commentaries.

WILEY-
BLACKWELL

For more information and to subscribe online, visit
wileyonlinelibrary.com/journal/infa

Society for Research in Child Development

IS Developmental Science

SRCD Exchange
exchange.srcd.org

SRCD's Social Media Site
Connect with your colleagues in the field!

Top 5 reasons why YOU need to be a SRCD member...

1. Play an Active Role & Broaden Your Horizons in the Society & the Field
2. Link, Share, & Connect with Policy Matters
3. Be "In the Know"
4. Opportunities to Apply for Grants & Awards
5. Many Member Extras

Join SRCD Today!
www.srcd.org

Questions regarding SRCD Membership contact: membership@srcd.org

558013

UNITED STATES POSTAL SERVICE®

Statement of Ownership, Management, and Circulation
(All Periodicals Publications Except Requester Publications)

1. Publication Title: Child Development

2. Publication Number: 8 8 9 _ 0 2 0

3. Filing Date: 10/1/10

4. Issue Frequency: Bi-monthly

5. Number of Issues Published Annually: 6

6. Annual Subscription Price: $609.00

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4®)*: Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

Contact Person: E. Schmidichen

Telephone *(Include area code)*: (201) 748-6346

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*: Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*: Wiley Subscription Services, Inc., 111 River Street, Hoboken, NJ 07030

Editor *(Name and complete mailing address)*: Jeffrey J Lockman, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118

Managing Editor *(Name and complete mailing address)*: Detra Davis, SRCD Executive Office, 2950 South State Street, Ste 401, Ann Arbor, MI 48104

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	**Complete Mailing Address**
Society for Research in Child Development	University of Michigan
	2950 South State Street, Ste 401
	Ann Arbor, MI 48104

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☑ None

Full Name	**Complete Mailing Address**

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☐ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

13. Publication Title: Child Development

14. Issue Date for Circulation Data: October 2010

15. **Extent and Nature of Circulation**			**Average No. Copies Each Issue During Preceding 12 Months**	**No. Copies of Single Issue Published Nearest to Filing Date**
a. Total Number of Copies *(Net press run)*			6104	5370
b. Paid Circulation *(By Mail and Outside the Mail)*	(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*	4122	4730
	(2)	Mailed In-County Paid Subscriptions Stated on PS Form 3541 *(Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)*	0	0
	(3)	Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	0	0
	(4)	Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)	0	0
c. Total Paid Distribution *(Sum of 15b (1), (2),(3), and (4))*			4122	4730
d. Free or Nominal Rate Distribution *(By Mail and Outside the Mail)*	(1)	Free or Nominal Rate Outside-County Copies Iincluded on PS Form 3541	92	89
	(2)	Free or Nominal Rate In-County Copies Included on PS Form 3541	0	0
	(3)	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)	0	0
	(4)	Free or Nominal Rate Distribution Outside the Mail *(Carriers or other means)*	0	0
e. Total Free or Nominal Rate Distribution *(Sum of 15d (1), (2), (3) and (4)*			92	89
f. Total Distribution *(Sum of 15c and 15e)* ▶			4214	4819
g. Copies not Distributed *(See Instructions to Publishers #4 (page #3))* ▶			1890	551
h. Total *(Sum of 15f and g)* ▶			6104	5370
i. Percent Paid *(15c divided by 15f times 100)* ▶			97.82	98.15

16. Publication of Statement of Ownership
☑ If the publication is a general publication, publication of this statement is required. Will be printed in the December 2010 issue of this publication.
☐ Publication not required

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Elizabeth Konkle, Associate Financial Manager

Date: 10/1/10

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and … are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

FIRST-YEAR MATERNAL EMPLOYMENT AND CHILD DEVELOPMENT IN THE FIRST 7 YEARS

- *Jeanne Brooks-Gunn, Wen-Jui Han, and Jane Waldfogel* (Serial No. 296, 2010).

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION

- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial No. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET

- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY

- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY

- *Mona El-Sheikh et al.* (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN

- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING

- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE

- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS

- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE

- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT

- *Michelle M. Chouinard* (Serial No. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS

- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM

- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997

- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD

- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS

- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT

- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES

- *Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES

- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR

- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

WILEY-
BLACKWELL